**THE APPLICATION OF HISTONE ANALYSIS FOR IDENTIFYING DISTAL REGULATORY ELEMENTS – REGULATION OF *FSHR***

BY
Elizabeth A. Dille

Submitted to the graduate degree program in Molecular and Integrative Physiology
and the Graduate Faculty of the University of Kansas Medical Center
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy.

_____
Leslie L. Heckert, Ph.D.
Chairperson

Committee Members:

_____
David F. Albertini, Ph.D.

_____
Vargheese M. Chennathukuzhi, Ph.D.

_____
Patrick E. Fields, Ph.D.

_____
Kenneth R. Peterson, Ph.D.

Date Defended: April 16, 2013

The Dissertation Committee for Elizabeth A. Dille certifies
that this is the approved version of the following dissertation:

**THE APPLICATION OF HISTONE ANALYSIS FOR IDENTIFYING DISTAL REGULATORY ELEMENTS – REGULATION OF *FSHR***

_____
Chairperson: Leslie L. Heckert, Ph.D.

Date Approved: April 19, 2013

## Abstract

Mechanisms that regulate gene expression are fundamental to many complex biological processes and disease states. Genome-wide approaches that combine chromatin-immunoprecipitation (ChIP) and next-generation sequencing have greatly advanced our understanding of chromatin structure and the role that histone modification plays in transcriptional regulation. In particular, these advances revealed important associations between functional, non-coding DNAs and specific histone modifications, which have been used technically to identify numerous distal regulatory elements and furthered our knowledge of transcriptional regulation and cell-specific gene regulation. *Fshr* is a gene expressed only in testicular Sertoli cells and ovarian granulosa cells and its expression is critical for proper gonad function and fertility. Importantly, underlying its exquisite cell-specificity is a transcriptional mechanism limited to only two cell types, which previous studies revealed was dependent on distal regulatory elements. To help identify these elements, we used ChIP, combined with next-generation sequencing, to globally map Histone 3 Lysine 4 tri-methylation (H3K4me3) in Sertoli and myoid cells. H3K4me3 is a post-translational histone modification known to associate with distal regulatory elements and promoter regions of actively transcribed genes. Analysis of H3K4me3 enrichment profiles identified a distal site 3' to *Fshr* that was specific to Sertoli cells. Transient transfection analysis indicated the region represses *Fshr* promoter activity and *in vitro* binding analysis revealed that GATA-4 and an unknown protein bound to the region, implicating them in cell-specific regulation of *Fshr*. These studies provide evidence that *Fshr* is regulated by a distal

regulatory element and have provided insight into the nature of these regulatory proteins. This work provides a database mapping H3K4me3 enrichment within Sertoli and peritubular myoid cells that can be used to identify new regulatory regions. Overall, these studies have furthered our knowledge of cell-specific gene regulation in Sertoli cells and provided new data that will lead to a better understanding of transcriptional regulation in Sertoli cells.

For Atropine.

# Acknowledgements

First I would like to thank my mentor, Dr. Leslie Heckert, for her input and guidance throughout my dissertation project. Her mentorship and guidance allowed me to work independently while still helping me develop the necessary skills needed for proper scientific investigation. She enabled me to expand on a project that has been a focus of her work for many years. Leslie also allowed me to participate in teaching outside of the laboratory environment, which improved my teaching skills and gave me the opportunity to explore multiple career choices.

I would like to thank each of my committee members for their guidance and input that helped me focus my dissertation. Drs. David Albertini, Vargheese Chennathukuzhi, Patrick Fields and Kenneth Peterson not only shaped my research, but also helped further my development as a scientist. Each personally invested time into my scientific growth and has contributed sound advice for the development of my professional career. I would also like to thank the Molecular and Integrative Physiology department and Chair, Dr. Paul Cheney, for all of their support.

Most of my laboratory work would not have been possible without help from Daren Rice. His expertise within the lab and through knowledge of molecular biology provided support for not only difficult experiments, but also for designing and implementing techniques. He is not only a work colleague, but has also become a good friend. Our numerous conversations about life made my daily lab experiences enjoyable.

I have made many friends during my time in Kansas City and each have contributed to my success in many ways. I would like to that Emily McDonald for all

of her support within and outside the lab. Your support definitely kept me going and I really miss our weekly 'coffee hour'. You were a blessing and a joy and I am sure our friendship will continue even though so many miles separate us.

Perhaps the strongest support of all came from my family. I am forever grateful for such a wonderful and supportive family. Both my mother, father and step-mother (Hazel, Tom and Corey) have continuously supported me throughout my college career and their confidence in my ability to succeed never wavered. Though separated by many miles, we have always maintained constant contact and I am truly thankful for all of their love and support. I want to thank my brother Tom for loving and accepting me for who I am and for believing in me.

I would also like to thank John Deuser all of his love and support. You have always had faith in me and for that I am thankful. You not only love me, but also loved my Atropine. You accepted both of us into your life, quirks and all. You not only allowed us into your life, but also shared all of your wonderful friends. As such I would like to thank the West Plaza Tomato Company for support, advice and a great place to decompress. You make life enjoyable and are a great reminder of what is really important in life – Great friends and good fun.

# Table of Contents

# List of Tables and Figures

## Tables

**Figures**

# Chapter 1

## Introduction

The writing in this chapter has been previously published in Biology of Reproduction.

Current concepts of follicle-stimulating hormone receptor gene regulation.
George JW, Dille EA, Heckert LL.
Biol Reprod. 2011 Jan;84(1):7-17.

## The Gonadotropin Hormones

Luteinizing hormone (LH) and follicle-stimulating hormone (FSH), are integral parts of the neural and endocrine interchange between the hypothalamus, pituitary, and gonads that controls steroid hormone synthesis and gamete production (reviewed in (Young 1995; Achermann and Jameson 1999; Plant 2008)). At the top of the network is gonadotropin releasing hormone (GnRH), which, once released from the hypothalamus, binds receptors on pituitary gonadotrophs and induces the synthesis and secretion of LH and the FSH. Once in circulation, LH and FSH finalize the communication by binding their receptors and transmitting signals to the gonads. These signals are at the hub of the regulatory network, relaying neuronal signals from the hypothalamus to the gonads and inducing feedback signals returned to the hypothalamus and pituitary. The receptors for FSH and LH, FSHR and LHR, reside on the surface of somatic cells in the gonads and are members of the Rhodopsin receptor family of G-protein coupled receptors, but unlike the other members, LHR and FSHR have extended NH2-terminal extracellular domains with numerous leucine-rich repeats that assist ligand specificity (Braun, Schofield et al. 1991; Dias, Cohen et al. 2002; Vassart, Pardo et al. 2004; Bogerd ; Lagerstrom and Schioth). FSH binding elicits several diverse signaling events, but the most characterized is that initiated by adenylyl cyclase, followed by induction of cAMP, PKA activation, and protein phosphorylation (Heindel, Rothenberg et al. 1975; Dorrington and Armstrong). FSH binding is also associated with an increase in intercellular calcium, activation of mitogen activated protein kinase (MAPK), and

stimulation of inositol triphosphate (IP3) (Flores, Veldhuis et al. 1990; Tena-Sempere, Manna et al. 1999; Seger, Hanoch et al. 2001).

Because the receptor is required for signal transduction, the original message relayed by FSH is delivered only to cells expressing FSHR and influenced by receptor levels. Thus, FSHR expression determines both the targets and extent of FSH action, ultimately directing hormone response to granulosa cells in the ovary and Sertoli cells in the testis (Heindel, Rothenberg et al. 1975). In the ovarian granulosa cells, temporal changes in FSH signaling regulate a number of transcriptional, metabolic, and hormonal activities that are important for the proliferation and differentiation events required for follicular growth and oocyte maturation (Abou-Issa and Reichert 1977; Peluso and Steger 1978; Grasso and Reichert 1990; Dunkel, Tilly et al. 1994; Rannikki, Zhang et al. 1995; Sairam, Jiang et al. 1996; Kumar, Wang et al. 1997; Simoni, Gromoll et al. 1997; Simoni, Gromoll et al. 1997; Huhtaniemi and Themmen 2005). In testicular Sertoli cells, the actions of FSH change with testis development [reviewed in (Kishi, Minegishi et al. 1998; Meachem, Ruwanpura et al. 2005)]. Initially, during the perinatal period, FSH induces Sertoli cell proliferation and establishes the final Sertoli cell number that will ultimately determine spermatogenic output, while later in development FSH stimulates Sertoli cell transcriptional and metabolic activities, which contribute to the hormonal and nutritional environment necessary for germ cell survival and development (Orth, Higginbotham et al. 1984; Russell and Griswold 1993; Spiteri-Grech, Weinbauer et al. 1993; Boitani, Stefanini et al. 1995; Meachem, McLachlan et al. 1996; Shetty, Marathe et al. 1996; Singh and Handelsman 1996; Ruwanpura, McLachlan et al.

2008). In both males and females, FSH induces hormonal signals that return to the pituitary and hypothalamus, as part of the feedback mechanism upholding the endocrine balance in the reproductive axis (Benson, Sorrentino et al. 1969; Moguilevsky, Libertun et al. 1970; Yen and Tsai 1971; Shahmanesh, Sedigh et al. 1980; Schwartz 1982; McNeilly, Souza et al. 2002).

## *FSHR* Expression

Expression of FSHR, both protein and mRNA, is remarkably limited with respect to its cellular profile, with Sertoli and granulosa cells by far the predominant expressing cell types (Ketelslegers and Catt 1974; Orth and Christensen 1977; Heckert and Griswold 1991; Dankbar, Brinkworth et al. 1995). *FSHR/Fshr* (Human/mouse gene) transcripts are first observed in embryonic gonads, around embryonic day 14.5 in males and 20.5 in females (Dankbar, Brinkworth et al. 1995; Rannikki, Zhang et al. 1995). These initial transcripts are incomplete and represent only the extracellular portion of the receptor, with full-length mRNA expressed several days later (Richards, Ireland et al. 1976; Sokka and Huhtaniemi 1990; Rannikki, Zhang et al. 1995). In the rodent ovary, Fshr expression coincides with primary follicle formation and follicular development through the pre-antral stage, with initial full-length transcripts and hormone binding observed shortly after birth (around postnatal day 3) and continuing to increase through day 21 (Sokka and Huhtaniemi 1990; Dunkel, Tilly et al. 1994; O'Shaughnessy, Marsh et al. 1994; Rannikki, Zhang et al. 1995; Drummond, Dyson et al. 1996). In the testis, full-length

FSHR mRNA initiates during fetal development (around embryonic day 16.5 in the rat) and expression is maintained throughout development and in the adult testis, (Steinberger, Thanki et al. 1974; Nimrod, Erickson et al. 1976; Sprengel, Braun et al. 1990; Camp, Rahal et al. 1991; Heckert and Griswold ; Oshaughnessy, Marsh et al. 1994; Dankbar, Brinkworth et al. 1995; Rannikki, Zhang et al. 1995). Once the spermatogenic cycle is initiated, FSHR levels change with the cycle, with levels highest at stages X-II and lowest at VI-VII (Heckert and Griswold ; Kliesch, Penttila et al. 1992; Heckert and Griswold ; Rannikko, Penttila et al. 1996). Several signals that regulate ovarian and testicular physiology also influence FSHR expression. In the ovary, FSHR is regulated by a combination of transcriptional and posttranscriptional mechanisms induced by FSH and activin, and indirectly by follistatin through its influence on activin (Knecht, Ranta et al. 1983; Woodruff, D'Agostino et al. 1988; Sanford and Batten ; Nakatani, Shimasaki et al. 1991; Themmen, Blok et al. 1991; Nakamura, Minegishi et al. 1993; Sites, Patterson et al. 1994; Minegishi, Tano et al. 1995; Tano, Minegishi et al. 1995; Tano, Minegishi et al. 1997). In the testis, FSHR is primarily regulated by its ligand, which decreases its expression through several mechanisms, including membrane receptor internalization, mRNA stability, and transcriptional regulation (Jahnsen, Gordeladze et al. 1980; Fletcher and Reichert 1984; Shimizu, Tsutsui et al. 1987; Themmen, Blok et al. 1991; Monaco, Foulkes et al. 1995; Viswanathan, Wood et al. 2009).

While FSHR expression is largely considered gonad-specific, there are several reports of its presence outside the ovary and testis; in particular, uterus, prostate, bone, and the ovarian surface epithelia (Zheng, Magid et al. 1996; Ben-Josef, Yang et

al. 1999; Mariani, Salvatori et al. 2006; Sun, Peng et al. 2006). In prostate and ovarian epithelial cells, FSH signaling is implicated in cell proliferation and tumor invasiveness in precancerous and malignant cells (Ben-Josef, Yang et al. 1999; Choi, Choi et al. 2004; Ji, Liu et al. 2004; Zhang, Chen et al. 2009). In bone, there is strong evidence linking FSH and FSHR to hypogonadal bone loss in women but there is still uncertainty as to the site of FSH action (Sun, Peng et al. 2006; Prior 2007; Zaidi, Blair et al. 2007; Zaidi, Zhu et al. 2007; Ritter, Thuering et al. 2008; Cannon, Cortez-Cooper et al. 2009; Zaidi, Blair et al. 2009; Robinson, Tourkova et al. 2010; Sun, Zhang et al. 2010). Consequently, it has been difficult to assess from the literature the degree to which normally differentiated cells, other than Sertoli and granulosa cells, actively transcribe *FSHR/Fshr*. However, this specificity can be more accurately appreciated by evaluating high-throughput expression data available through the public domain. As an example, data generated as part of the mouse transcriptome project, derived by massively parallel signature sequencing of over 30 tissues, showed *Fshr* mRNA present in the ovary at much higher levels than in the testis (Figure 1-1, from GEO Record GDS868). However, the expression signatures were from adult tissues and therefore the testis profile reflects a pool of expressing Sertoli cells that is significantly diluted by an extended germ cell population, while not such dilution effect occurs in the ovary. Thus, on the cellular level, expression is expected to be more equitable. What is most striking in these findings is that, with this highly sensitive method, no other tissues showed detectable *FSHR/Fshr* expression. What is the mechanism that is responsible for this remarkable cell-specificity? Currently, it is unknown but, as reviewed below,

the evidence indicates that *FSHR* expression is directed by factors that control its transcription through elements located at significant distance from the gene itself.

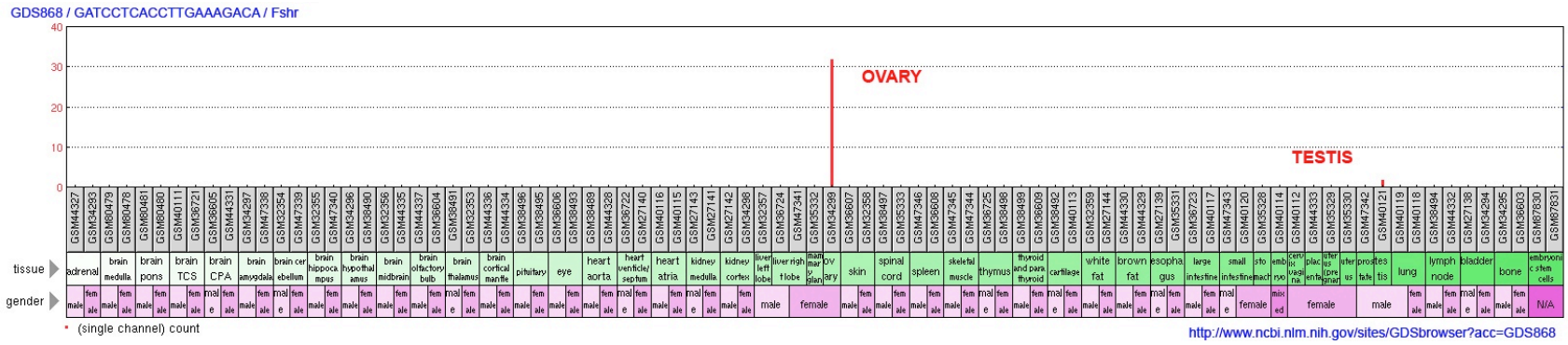



**Figure 1-1. Tissue expression profile of mouse *Fshr*.** Gene query results from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) derived using terms for FSHR and dataset record GDS868 (Edgar, Domrachev et al. 2002; Barrett, Troup et al. 2007; Barrett, Troup et al. 2009). Data represents large scale transcriptome analysis of tissues from the C57BL/6J mouse strain, using massively parallel signature sequencing technology and are accessible through GEO Series accession number GSE1581 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1581).

## *FSHR/Fshr* Transcription

Transcription of *FSHR/Fshr* contributes to receptor levels and directs its cell-specificity, indicating components of the underlying mechanisms are important for both FSH responsiveness and target cell identity. Our current understanding of FSHR/Fshr transcription is derived largely from studies on the rat, murine, ovine, and human FSHR/Fshr genes, which focused on characterization of the 5' flanking region. The resulting data provided significant insight, revealing both similarities and differences between promoters of the four species. Since most promoter characteristics were detailed in a previous review, the discussion here will be limited to its prominent features (Heckert 2005).

The accumulated information represents primarily transient transfection and DNA/protein binding results from Sertoli and granulosa cells that identified regulatory elements and their associated binding proteins within promoters represented by various species and lengths and revealed both similarities and differences in promoter function (Huhtaniemi, Eskola et al. 1992; Gromoll, Dankbar et al. 1994; Goetz, Lloyd et al. 1996; Heckert, Daggett et al. 1998; Heckert, Sawadogo et al. 2000; Heckert 2001; Kim and Griswold 2001; Xing and Sairam 2001; Xing, Danilovich et al. 2002; Xing and Sairam 2002; Xing and Sairam 2002). Promoter sequence comparisons between several species showed significant conservation of approximately 1000bp 5' to the translational start, a reference point used to avoid uncertainty in transcriptional start sites for some species (Huhtaniemi, Eskola et al. 1992; Gromoll, Dankbar et al. 1994; Sairam and Subbarayan 1997; Tena-Sempere, Manna et al. 1999; Xing and Sairam 2001; Heckert and Griswold 2002; Heckert

2005). While variations in methodology and promoter context make it hard to assess relevance to regulatory sequences identified in a single species or study, results representing the rat, sheep, and human promoters identified a common E box element (Goetz, Lloyd et al. 1996; Heckert, Daggett et al. 1998; Findlay and Drummond 1999; Xing and Sairam 2001; Xing, Danilovich et al. 2002; Putowski, Schillings et al. 2004). Located just 5' to the transcriptional start sites, the element contributes significantly to promoter activity and provides a common mechanistic theme that features the E box and its cognate binding factors upstream stimulatory factor 1 (USF1) and 2 (USF2) (Figure 1-2). The USF proteins are members of the helix-loop-helix family and form both homo- and heterodimers (Shieh, Sparkes et al. 1993; Viollet, Lefrancois-Martinez et al. 1996; Rodriguez, Girones et al. 2003). There is considerable evidence, both *in vitro* and *in vivo*, documenting the importance of the E box and the USF proteins for *Fshr* promoter activity (Goetz, Lloyd et al. 1996; Heckert, Daggett et al. 1998; Findlay and Drummond 1999; Heckert, Sawadogo et al. 2000; Griswold and Kim 2001; Heckert 2001). Studies on USF1 and USF2 also showed differential effects on *Fshr* expression in testis and ovary and, in Sertoli cells, FSH modulates their activity via the inhibitory protein ID2 (Deoxyribonucleic Acid Binding/Differentiation-2), which likely explains the corresponding change in *Fshr* mRNA levels (Hermann, Hornbaker et al. 2008; Viswanathan, Wood et al. 2009). Other proteins noted for regulating the *Fshr* promoter include steroidogenic factor-1 (SF-1), SMAD3, E2F, GATA-1, and ETS, several of which bind sequences that, when mutated, reduce promoter activity

(Heckert 2001; Kim and Griswold 2001; Levallet, Koskimies et al. 2001; Gong and McGee 2009; Brune, Adams et al. 2010).

Despite numerous studies on the promoter, no mechanism evolved that could explain its cell specificity. This deficiency in promoter specificity was also demonstrated by studies that evaluated 16 distinct transgenic mouse lines for cell-specific expression of reporters directed by either 5.0kbp or 143bp (8 lines each) of rat *Fshr* promoter sequence, none of which showed Sertoli or granulosa cell specific expression (Heckert, Sawadogo et al. 2000). Similar finding were reported for 1.5kbp of the human promoter (Nordhoff, Gromoll et al. 2003). The recognition that sequences beyond the 5.0kbp promoter were required for expression, together with emerging data on regulatory elements that act from distal positions, led to studies using yeast artificial chromosomes (YACs) as transgenes, in order to define the region required for FSHR specificity. While YAC transgenic data is limited to a single mouse line, which harbors the entire rat *Fshr* gene plus sequences that span >50kbp 5' and >30kbp 3', its failed detection in Sertoli and granulosa cells supports Fshr distal regulation and suggests important sequences extend well beyond the gene (Hermann, Hornbaker et al. 2007).

With the evidence that proper expression of FSHR/Fshr requires contributions from regulatory elements located outside the promoter region, most likely at significant distances, the challenge became identifying these sites within a vast amount of potential sequence. Initially, with only a small amount of available sequence, this was tackled using conventional DNase I hypersensitivity mapping to identify regions of accessible chromatin and thus potential sites of regulatory

importance (Hermann and Heckert 2005). This revealed four hypersensitive sites located within a 45 kb region surrounding the first exon, three of which showed significant sequence conservation, which is also a predictive feature of important regulatory elements (Hermann and Heckert 2005). One of these, DHS#3, was located approximately 4kb downstream and showed much greater sensitivity to DNase I in non-expressing cells (myoid) than expressing cells (Sertoli), suggesting an association with gene silencing (Figure V-2). Further functional characterization, using transient transfections and *in vitro* and *in vivo* binding assays, revealed important elements that attenuated gene expression and their cognate binding proteins, OCT1, GATA4, and GATA1. The studies also implicated OCT1 in selective binding to this element in order to maintain its silent state in non-expressing cells (Hermann and Heckert 2005). While the approach proved to be a valid means to identify important regulatory elements, it was evident that scanning extensive regions of the genome without better knowledge of its DNA content or genomic landscape was impractical. This has since been remedied by the infusion of genomics data that brought, not only new sequence information, but also a wealth of insight on the *FSHR/Fshr* gene and it residing landscape.

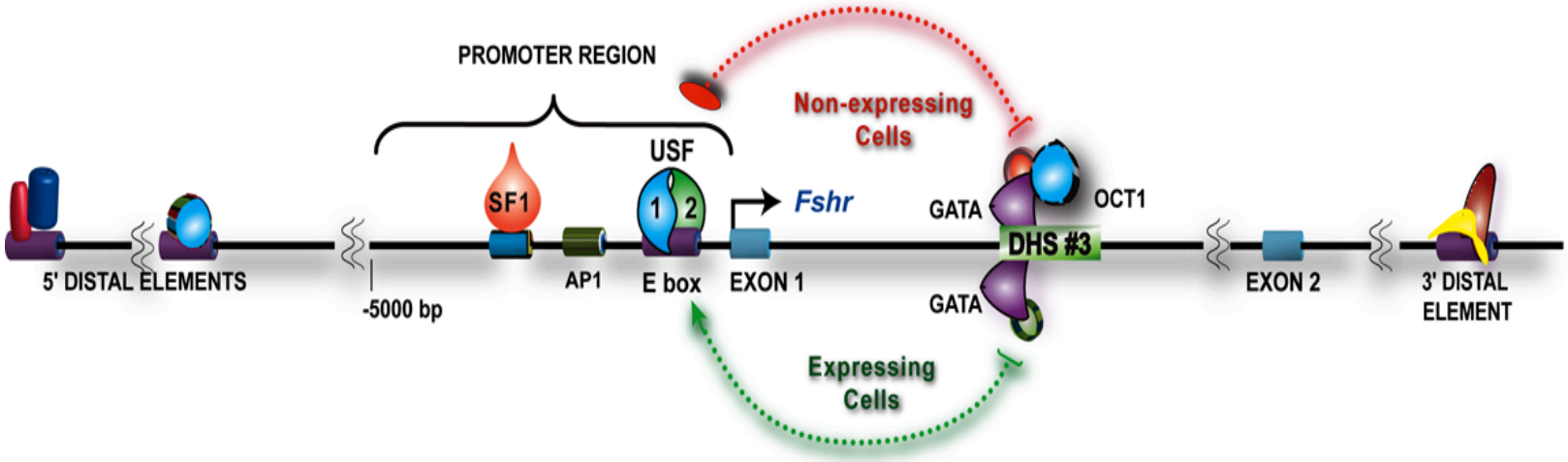



**Figure 1-2. Summary of *FSHR/Fshr* transcriptional regulation.** Identified elements within the *FSHR/Fshr* promoter region are depicted (rectangles) and shapes above the elements indicated the identified binding proteins SF1 (steroidogenic factor 1) and USF (Upstream stimulatory factor). The relative binding affinity noted in the ovary and testis for USF1 and USF1 is indicated below. DHS#3 noted in the first exon is bound by a GATA-containing complex that differs with respect to the presence of OCT1, which also correlates with the silencing activity observed in non-expressing cells. The presence of distal regulatory elements as predicted by several studies, is also indicated.

## The *FSHR* Gene

The initial characterization of the *FHSR* gene in 1992 revealed a pronounced structural similarity to the LHR gene that suggested the two evolved through duplication of a common ancestral gene (Heckert, Daley et al. 1992). Similarities between *FSHR/Fshr* and *LHR/Lhr* and the genes for other G-protein coupled receptors (GPCR) also suggested the predecessor for *FSHR/Fshr* and *LHR/Lhr* was formed by combining a common GPCR ancestral gene, encoding the characteristic transmembrane and intracellular domains, with multiple repeated exons derived from tandem duplications of a module for a leucine-rich motif, a featured attribute of the extracellular domain of glycoprotein hormone receptors (Heckert, Daley et al. 1992). Formation of the gonadotropin receptor genes through tandem duplication of an ancestral gene was further substantiated by recent genomics data that show they are arranged in tandem on chromosomes from nearly all annotated Tetrapoda genomes (Heckert, Daley et al. 1992; Montgomery, Tate et al. 1995; Chauvigne, Tingaud-Sequeira et al. 2010). While the two ancestral descendants, *FSHR/Fshr* and *LHR/Lhr,* differ by one exon (10 for FSHR/Fshr and 11 for *LHR/Lhr*), the coding scheme is largely the same, with the carboxy-terminal, transmembrane-intracellular domain encoded by the last exon and the amino-terminal, extracellular domain by all preceding exons (Koo, Ji et al. 1991; Tsai-Morris, Buczko et al. 1991; Heckert, Daley et al. 1992). Exons that partition the receptors' extracellular domains also share a repeated structure that delineates seven of the leucine-rich motifs into exons 2-8 and two into exon 9 (Figure 1-3).

However, despite knowing the structure of *FSHR/Fshr* for nearly two decades, it wasn't until the various genome-sequencing projects greatly expanded the available sequence data that there was accurate knowledge of its size or chromosome habitat. Now, *FSHR/Fshr* chromosome locations and annotated sequences are reported for more than 40 vertebrate species through the University of California, Santa Cruz Genome Browser (http://genome.ucsc.edu/) and similar web sites. In human, rat, and mouse *FSHR/Fshr* are located on chromosomes 2, 6 and 17, respectively, and spans roughly 200kb, a size much larger than originally predicted (Heckert, Daley et al. 1992; Huhtaniemi, Eskola et al. 1992; Rousseau-Merck, Atger et al. 1993; Gromoll, Ried et al. 1994; Rhead, Karolchik et al. 2010) (Figure 1-3). This wealth of sequence information also disclosed several defining features of *FSHR/Fshr* and its surrounding neighborhood that indicates its associated regulatory environment is strongly influenced by evolutionary constraint that retains regulatory sequences directing distally-located genes (Hermann and Heckert 2005; Hermann, Hornbaker et al. 2007). This includes the tandem placement of *FSHR/Fshr* between *NRXN1/Nrxn1*, it closest 5' neighbor, and *LHR/Lhr*, its closest 3' neighbor, to form a highly conserved syntenic block, which was identified by several studies to indicate evolutionary constraints intended to preserve the alignment between noncoding sequences and their target genes (Hermann and Heckert 2005; Engstrom, Ho Sui et al. 2007; Kikuta, Laplante et al. 2007; Akalin, Fredman et al. 2009).

Another notable feature is the large intergenic distances between *FSHR/Fshr* and its adjoining neighbors, which span more than 750kbp on its 5' side and 200kbp

on its 3‘ side (Hermann and Heckert 2005). The uninhabited region between *NRXN1/Nrxn1* and *FSHR/Fshr* is characteristic of a gene desert, or long genomic stretch devoid of protein-coding sequences or other obvious biological functions. Such regions, when located within a conserved syntenic block, are associated with areas of enhanced sequence conservation, suggesting the content was under evolutionary pressure to retain functional and contextual information of resident elements (Venter, Adams et al. 2001; Ovcharenko, Loots et al. 2005; Akalin, Fredman et al. 2009). While evidence shows not all gene deserts have measurable activity, numerous risk loci and regulatory elements are documented within these regions, confirming their importance to the genome (Lodder, Eussen et al. 2009; Xu, Tsumagari et al. 2009; Kiltie 2010). Comparative sequence analysis has been consistently used in genome-wide studies to evaluate evolutionary constraint as a means to understanding genome structure, biological function, and evolution (Waterston, Lindblad-Toh et al. 2002). The initial whole-genome sequence comparisons between mouse and human provided considerable insight and estimated 5% of the human genome was conserved over 70-100 million years (Lander, Linton et al. 2001; Venter, Adams et al. 2001; Dermitzakis, Reymond et al. 2002; Mural, Adams et al. 2002; Waterston, Lindblad-Toh et al. 2002; Pennacchio 2003). What was remarkable in this finding was that only 1/3 of the conserved sequences was located in coding regions. Thus, the genome’s non-coding sector represents the largest portion under evolutionary selection, which suggested there is considerable functional information within conserved non-coding sequences, aka

evolutionary conserved regions (ECRs) (Waterston, Lindblad-Toh et al. 2002; Dermitzakis, Reymond et al. 2005).

Expansion of genome sequence data and specie representation added significantly to the power of cross-species sequence comparison and the genomic landscape and sites of potential biological functions, which led to identification of many non-coding sequences involved in gene regulation, including enhancers, insulators, silencers and matrix attachment regions (Glazko, Koonin et al. 2003; Koonin 2003; Nobrega, Ovcharenko et al. 2003; de la Calle-Mustienes, Feijoo et al. 2005; Woolfe, Goodson et al. 2005; Pennacchio, Ahituv et al. 2006; Prabhakar, Poulin et al. 2006; Visel, Prabhakar et al. 2008). A connection between sequence conservation and regulatory sequences was nicely illustrated in a recent study that used chromatin immunoprecipitation (ChIP) and massively parallel sequencing to map DNA sites linked to p300, a co-regulator for many transcription factors (Visel, Blow et al. 2009). Results from mouse embryonic tissues (forebrain, midbrain and limb) showed p300 highly enriched at sites containing conserved non-coding sequence and, of those evaluated, nearly all were functional. The study also demonstrated that most p300-associated sites/enhancers were located at least 10kbp from a potential target gene, which suggests distal elements are commonly involved in transcriptional control. Direct evidence for long-distance gene regulation is found in numerous studies that demonstrate transcriptional effects from distant sequences on specific target genes, such as those for β-globin, IFγ, SOX9, GATA3, FGF4, IL-10, and CD69 (Martinez-Jimenez, Gomez-Lechon et al. 2005; Bejerano, Lowe et al. 2006; Schoenborn, Dorschner et al. 2007; Vazquez, Laguna et

al. 2009). The significance of non-coding regulatory sequences to human disorders has also gained recognition with reports linking them to diseases and developmental disorders, as noted in cases of x-linked deafness, preaxial polydactyl, campomelic dysplasia, sex reversal, postaxial polydactyly (de Kok, Merkx et al. 1995; Bishop, Whitworth et al. 2000; Jamieson, Perveen et al. 2002; Epstein 2009; Lodder, Eussen et al. 2009). In the case of preaxial polydacyl, the reported mutation was located within a regulatory element approximately 1Mbp from its target gene that emphasizes not only the remarkable linear distance in which the transcriptional signal passes but also the importance of nuclear architecture in positioning regions that collaborate in the transcriptional signal (reviewed in (Schneider and Grosschedl 2007)).

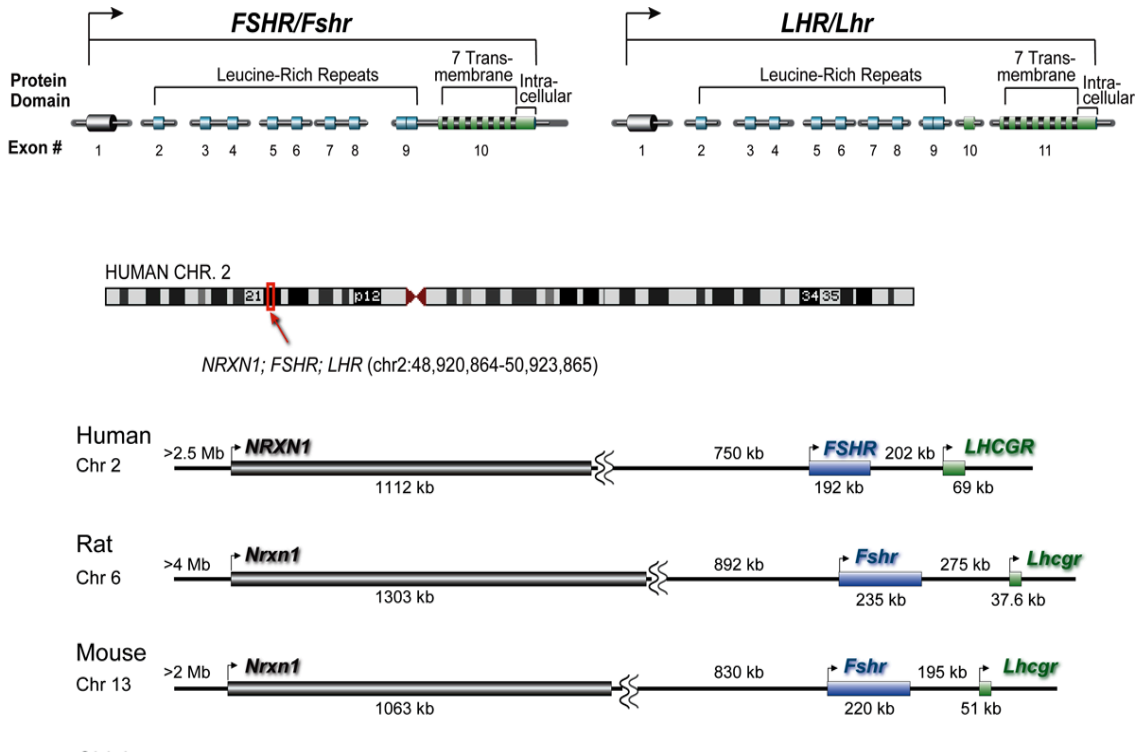



**Figure 1-3. Organization of *FSHR/Fshr* and its genomic environment.** Top, structural *FSHR/Fshr* and *LHCGR/Lhcgr*, depicting the exon distribution of the receptor's protein domains. Middle, position of the *NRXN1;FSHR;LH/CGR* syntenic region on human chromosome 2 (red box indicated by arrow). Bottom, syntenic region of *NRXN1/Nrxn1, FSHR/Fshr*, and *LHCGR/Lhcgr* from human, rat, mouse, and chicken, with sizes indicated for the genes and intervening regions. Exons (top) and genes (bottom) are indicated by rectangles, intervening regions by lines, and transcriptional direction by arrows.

## The Search for New Regulatory Elements

Identification of regulatory regions is clearly facilitated by predictions based on sequence conservation. However, the evidence is also clear that sequence conservation does not detect all regulatory elements nor provide any assurance that the predicted sites are functional, and thus, most effective when used in conjunction with other corroborating techniques, particularly ones that can assess chromatin changes linked to transcriptional activity (Nobrega, Ovcharenko et al. 2003; Dermitzakis, Reymond et al. 2005; Rizzolio, Bione et al. 2008; D'Haene, Attanasio et al. 2009; Liska, Snajdr et al. 2009; Vazquez, Laguna et al. 2009; Visel, Zhu et al. 2010). In studies to identify regulatory sequences that direct *FSHR/Fshr* transcription, results from transgenic and promoter studies shifted the experimental focus away from the promoter to the region encompassing all of *FSHR/Fshr* and its adjoining intergenic regions, which required a new set of tools and resources that fortunately evolved from the collection of genome sequences and efforts to understand their content. Initial reports on the *FSHR/Fshr* conservation profile compared human and rat sequences by direct pairwise comparison and analysis of precompiled LAGAN alignments through the web-based VISTA genome browser (Hermann and Heckert 2005). This revealed over 150 conserved sites, which, when matched together with DNase I hypersensitivity data, was instrumental in the identification of an important silencing region in the first intron (discussed above). However, it was also evident that greater constraints were required to improve functional prediction and therefore, once more distant genomes (e.g. chicken) were available and included in computations, evolutionary conserved

regions (ECRs) with greater predictive power could be distinguished and seven of the most conserved were selected for functional testing by transient transfection (Hermann, Hornbaker et al. 2007). With continuous enhancements in genome data and resources that improve regulatory element prediction, the number of selected sites has grown to more than twenty, which includes the original seven ECRs. These sequences were identified using the human genome and the UCSC Genome Browser (http://genome.ucsc.edu/) to reveal highly conserved (vertebrate consensus and net alignment with chicken as the target), non-coding sequences with a significant regulatory potential score (7X regulatory potential; (Waterston, Lindblad-Toh et al. 2002; Consortium 2004; Gibbs, Weinstock et al. 2004; Havlak, Chen et al. 2004; Kolbe, Taylor et al. 2004; King, Taylor et al. 2007)). Figure 1-4 provides an example of the *FSHR* locus spanning from *LHCGR* to *NRXN1* that was modified from the UCSC Genome Browser results to show the top predicted sites (marked ECRs) for *FSHR/Fshr* regulatory sequences and the key features used in their selection.

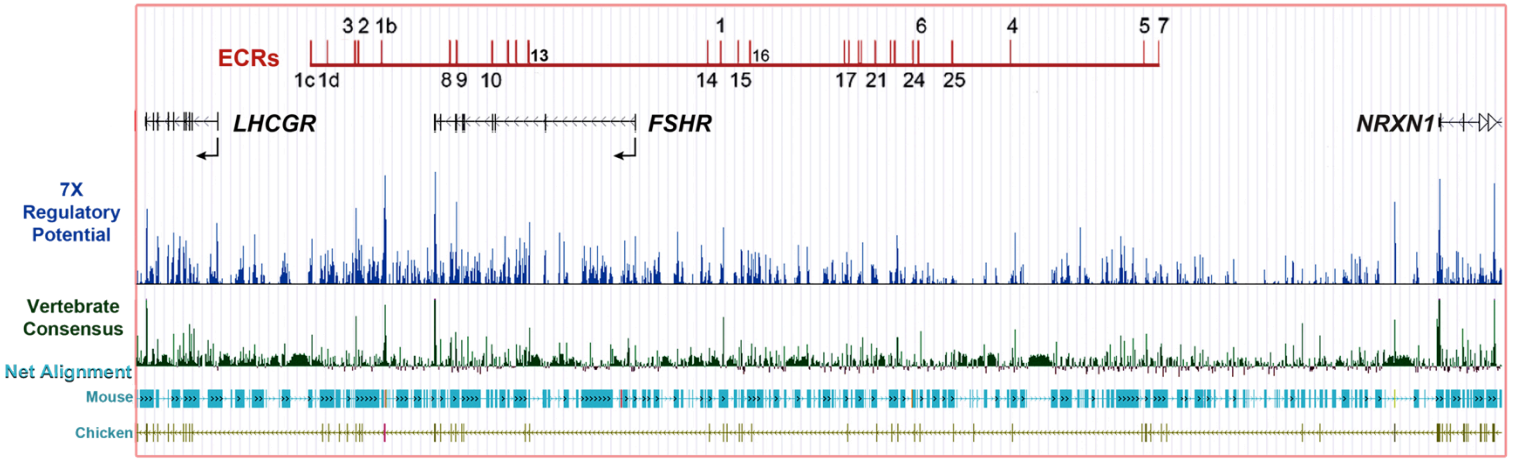



**Figure 1-4. Annotated region of human *FSHR* and neighboring genes *NRXN1* and *LHCGR* from the UCSC Genome Browser** (http://genome.ucsc.edu/). With human as the reference genome, noncoding evolutionary conserved regions (ECRs, top) were selected for their potential regulatory activity using sequence conservation, both vertebrate consensus and net alignment (chicken as the target), and 7X regulatory potential. Vertebrate consensus represents alignments of 44 vertebrate species generated using multiz and other tools and measurements of evolutionary conservation determined by phastCons and phyloP (Siepel, Bejerano et al. 2005; Pollard, Hubisz et al. 2010). The 7X regulatory potential track represents scores computed from human, chimpanzee, macaque, mouse, rat, dog, and cow by comparing short alignment pattern frequencies between known regulatory elements and neutral DNA (Waterston, Lindblad-Toh et al. 2002; Consortium 2004; Gibbs, Weinstock et al. 2004; Havlak, Chen et al. 2004; Kolbe, Taylor et al. 2004; King, Taylor et al. 2007). The net alignment tracks show the best mouse/human or chicken/human chains for every part of the human genome, with ungapped alignments represented by boxes and gaps by lines. Annotated sequences are from the chicken May 2006 (WUGSC 2.1/galGal3) (galGal3) assembly, mouse July 2007 (NCBI37/mm9) (mm9) assembly, and human March 2006 (NCBI36/hg18) assembly.

## Summary

The evidence to date shows that sequences directing *FSHR/Fshr* expression lie far from the start of transcription in a regulatory environment without defined boundaries, which complicates their detection using standard molecular approaches. While computational genomics has helped narrow the search, limitations due to false positives and undetected sequences caution its use without additional methods to substantiate the data (Giresi, Kim et al. 2007). Fortunately, many technologies have adapted to the influx of sequence data by developing high-throughput and genome-wide strategies. Two such strategies offer considerable promise for regulatory element identification and that the *FSHR/Fshr* transcriptional mechanism is within reach. Both strategies reveal chromatin signatures featured in regulatory sequences; one identifies sequences bound to modified histones linked to transcriptional activity by chromatin immunoprecipitation, the other identifies open regions of chromatin, similar to DNase I hypersensitivity, by formaldehyde associated identification of regulatory elements (Wu, Smith et al. 2006; McGaughey, Stine et al. 2009). Implementation of such strategies together with comparative genomics will significantly enhance the probability of relevant sequence identification and the mechanistic understanding of the regulatory landscape. When used with high-throughput strategies, i.e. high-density tiling arrays and next generation sequencing, to canvass the genome without the bias of conservation, additional insight is likely on mechanisms that employ non-conserved regulatory elements and possible contributions to species-specific regulatory features (Roh, Wei et al. 2007; Chen, Lin et al. 2008).

# Chapter 2

## Mapping Post-translational Histone Modifications in P15 Sprague Dawley Sertoli and Myoid Cells

## Abstract

To advance our understanding of genome function and gene regulation in Sertoli cells, we employed Chromatin Immunoprecipitation combined with next-generation sequencing (ChIP-Seq) to globally map the histone landscape in Sertoli and myoid cells of the day 15 Sprague Dawley rat. Three histone modifications were selected for analysis based on their ability to identify active genomic regions, active enhancer/promoter regions and repressed genomic regions - Histone 4 Lysine acetylation (K5, 8, 12 and 16), Histone 3 Lysine 4 tri-methylation and Histone 3 Lysine 27 tri-methylation (respectively). Raw ChIP-Seq sequence tags were processed and aligned to the *rattus norvegicus* genome (assembly: rn5 March 2012) using the Burrows-Wheeler Aligner. Datasets contain alignment profiles of the processed sequence tags generated by Illumina Sequencing and can be viewed using the UCSC Genome Browser. Overall, 45-80% of the sequence tags aligned to the genome except for the SC-H4ac and SC-HeK27me3 samples, which aligned poorly and were not used for further analysis. We have generated two reference databases that map the H3K4me3 profile across the *rattus norvigicus* genome in P15 Sertoli and myoid cells using two genome assemblies – a masked and unmasked genome. These databases can be used for peak finding analysis, identificaiton of regulatory regions and can be manually evaluated to identify H3K4me3 enrichment at specific gene loci. Use of these databases will further our understanding of the regulatory landscape in Sertoli cells and provide new insite into the transcriptional regulation within these cells.

# Introduction

The objective of this project is to elucidate transcriptional mechanisms relevant to Sertoli cell function by identifying new *cis*-regulatory elements. To obtain this objective, two testicular cell types will be included in our genome-wide analysis of histone modifications – Sertoli and peritubular myoid cells. These two cell types differ in their physiologic function and will provide a contrasting view of the histone landscape at different loci within the genome. Comparison of the histone landscape between these cell types will lead to the characterization of both common and cell-type specific regulatory regions and thus further our understanding of transcriptional regulation in these two cell types.

At P15, Sertoli cells are undergoing several physiologic changes. First, the Sertoli cells have ceased mitotic activity (Steinberger and Steinberger 1971; Russell, de Franca et al. 1995), and are entering their final stages of differentiation. Thus, the final number of Sertoli cells that will populate the adult testis is determined (Steinberger and Steinberger 1971). Aside from diminished mitotic activity, the Sertoli cells are also forming cell-to-cell junctions with each other, which leads to the formation of the blood-testis barrier. These intracellular junctions also act to compartmentalize developing germ cells throughout each stage of maturation, ensuring the proper niche for development (Ryser, Glauser et al. 2011).

The peritubular myoid cells form a single layer of cells that lye just outside the basal lamina of the seminiferous tubules in rodents (Bressler and Ross 1973). These cells are similar to smooth muscle cells and are believed to play an important role in the rhythmic contraction of the tubules (Leeson and Forman 1981). During

testis formation, the Sertoli cells and peritubular myoid cells work in concert with each other to establish the basil lamina, which is the main physical structure forming the tubule (Tung, Skinner et al. 1984). Aside from a structural and physical role in testis function, myoid cells interact with Sertoli cells in a paracrine manor to regulate Sertoli cell function (Raychoudhury, Thompson et al. 1993; Verhoeven, Hoeben et al. 2000). Identifying new *cis*-regulatory elements within these two cell types will further our understanding of testis development, cell-specific gene regulation, and contribute to our overall understanding of gene regulation within the testis.

Chromatin Immunoprecipitation (ChIP) provides a powerful tool that gives researchers the ability to understand protein/DNA association *in vivo* on a genome wide scale. ChIP is a powerful technique used to identify proteins bound to DNA *in vivo* and has provided enormous insight into chromatin structure. For ChIP analysis, nuclear proteins are cross-linked to DNA and immunoprecipitated using an antibody of interest. The resulting immunoprecipitated DNA pool can be analyzed by PCR to identify specific targeted regions or analyzed with next-generation sequencing to provide a genome-wide view.

The current study focused on identifying enrichment patterns of three well-studied histone modifications in Sertoli and myoid cells. We selected Histone 4 Lysine 5, 8, 12 and 16 acetylation (H4ac), Histone 3 Lysine 4 tri-methylation (H3K4me3) and Histone 3 Lysine 27 tri-methylation (H3K27me3) to identify euchromatic regions, active promoters and *cis*-regulatory elements and repressed regions (respectively).

The antibody selected for identification of H4ac recognizes chromatin regions with at least one, and possibly more, modification/s to Lysine 5, 8, 12 and 16. This mark was selected for its known association with active genes (Roh, Ngau et al. 2004; McCann, Muro et al. 2007; Ramsey, Knijnenburg et al. 2010). Previous studies have shown that histone acetylation is associated with larger chromatin regions that contain actively transcribed genes (reviewed in (Calestagne-Morelli and Ausio 2006)). Within these genes, acetylation is often found within the core promoter and gene, and we would like to determine if H4ac is correlated with actively transcribed promoters and genes within our cell populations.

Histone 3 Lysine 4 tri-methylation is a mark often associated with transcriptionally active promoter regions and enhancer regions (Bernstein, Kamal et al. 2005; Kim, Barrera et al. 2005; Heintzman, Hon et al. 2009). We selected this histone mark for two reasons: 1) To determine if this mark can be associated with active promoters in Sertoli/Spermatogonia cells, and 2) to determine if this mark can identify enhancer regions within our cell populations (discussed further in Chapter 3). As genome wide studies have identified peaks of H3K4me3 surrounding transcriptional start sites of active genes in multiple cell types (Schmid and Bucher 2007; Heintzman, Hon et al. 2009), we believe this mark will also be associated with active promoters within testis cell populations.

Histone 3 Lysine 27 tri-methylation is a mark associated with transcriptionally silent regions (Young, Willson et al. 2011). The current understanding of this methylation mark indicates that a broad region of H3K27me3 is often found at genes that are actively silenced (Pauler, Sloane et al. 2009).

Recently, H3K27me3 was identified within the promoter regions of transcriptionally active genes in association with H3K4me3 (Dahl, Reiner et al. 2010; McEwen and Ferguson-Smith 2010). It is believed that this active, yet repressive chromatin state acts to poise genes for activation. It is most commonly associated with the promoter region of genes that encode transcription factors that are critical for development (Bernstein, Mikkelsen et al. 2006; Young, Willson et al. 2011). For our study, this mark will serve to identify silenced regions and in conjunction with H3K4me3 identify poised genes.

The Sertoli and myoid cell enrichment profiles generated using these modifications display all of the sequence tags that aligned to the *Rattus norvegicus* (rn5, March 2012) genome in two formats, Masked and Unmasked. The sequence tags were aligned to a complete genome (Unmasked) and a genome that did not contain repeats (Masked). The resulting databases are searchable by chromosome location or gene name and can be used for peak finding analysis (see Chapter 3). All but two of our ChIP-Seq samples generated good quality genome-wide maps identifying cell-specific enrichment of each histone modification tested. The Sertoli cell samples for H4ac and H3K27me3 did not align well to the reference genome and will need re-sequenced before further analysis can be performed. We did obtain good-quality alignments for both Sertoli and myoid cell H3K4me3 and myoid cell H4ac and H3K27me3.

## Materials and Methods

**Animal Use.** All experiments using animals were approved by the Institutional Animal Care and Use Committee of the University of Kansas Medical Center (Protocol 2009-1850), and performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

**Sertoli Cell and Myoid Cell Preparations.** Sertoli and myoid cells were prepared as previously described with slight modification (Steinberger, Heindel et al. 1975; Karl and Griswold 1990). Briefly, tubules were isolated from P15 Sprague Dawley rats. Tubules were processed through a series of digestions releasing the myoid cells which, following centrifugation, remain in the supernatant while Sertoli cells and germ cells are collected in the pellet. During the isolation, myoid cells were removed from the tubules prior to complete digestion and subsequent release of the Sertoli and germ cells. The majority of germ cells remaining in the culture after digestion are spermatogonia as more advanced stages of germ cells are removed during tubule digestion. Sertoli cells were used fresh for ChIP assays and thus contain a subpopulation of spermatogonia. Myoid cells were cultured in Ham's F12 (Cellgro, Corning, NY -#10-080-CV), 10% FBS, 1.5mM HEPES (Sigma-Aldrich, St. Louis, MO - H0887) and 1% Pen-Strep (Sigma-Aldrich, St. Louis, MO - P0781). All cell culture was at 37°C in 5% $CO_2$. Sertoli cells were either used directly (Chromatin Immunoprecipitation) or cultured for two days and shocked with 10 mM Tris pH 7.4 for 2-4 minutes to remove germ cells. Myoid cells were cultured and passaged two times before use.

**Chromatin Preparations.** Chromatin was prepared as previously described (Hiroi, Christenson et al. 2004). Prior to cross linking, Sertoli and myoid cells were counted to establish cell numbers for digestion. To cross-link chromatin, primary rat Sertoli and myoid cells were and incubated in 30 mL cell media containing 1% Formaldehyde for 10 minutes at room temperature. Cross-linking was stopped by the addition of 125mM. Cells were pelleted and washed two times in 10 mL ice-cold PBS containing freshly prepared Protease Inhibitors (Roche, South San Francisco, CA-#, South San Francisco). Cell pellets were washed three times in 20 mL ice-cold MC lysis buffer (10mM Tris-HCl pH 7.5, 10mM NaCl, 3mM $MgCl_2$, 0.5% NP-40) and incubated on ice for 10 minutes during the third wash. After the final wash was complete the cell pellets were vortexed (setting 3) for three minutes. The cells were pelleted, the supernatant removed and the pellet frozen by immersion in liquid nitrogen. The pellets were then thawed in an ice-water bath. Cells were resuspended to 4 x $10^7$ cells/mL in MNase buffer (10mM Tris-HCl pH 7.5, 10mM NaCl, 3mM $MgCl_2$, 1mM $CaCl_2$, 4% NP-40) and disrupted using 10 strokes with a Type B pestle in an ice-cold Dounce homogenizer to release chromatin. Chromatin suspensions were then diluted with cold MNase buffer to a final concentration of 2 x $10^7$ cells/mL. Micrococcal nuclease (New England BioLabs M0247S) was added (120U per 2 x $10^6$ cells) and extracts were incubated for 30 minutes at 37°C. Nuclease digestion was stopped by the addition of (final concentration): 6mM EGTA, 1x protease inhibitors (Roche, South San Francisco, CA), 10% SDS and 200mM NaCl. Samples were aliquoted to 1 x $10^7$ cells (~500 μl) and stored at -80°C.

**Chromatin Immunoprecipitation**. Rat Sertoli and myoid chromatin was used for ChIP at an equivalent of 1 x $10^7$ cells per immunoprecipitation sample. Chromatin samples were thawed on ice and pre-cleared by the addition of 75 μl of a 50% slurry of pre-blocked protein A-agarose beads (Millipore 16-157) for one hour at 4°C with rotation. Agarose beads were then pelleted by centrifugation for one minute at 3900 x g at 4°C and the supernatants transferred to new tubes. Five micrograms of primary antibody was added to the supernatants and the samples incubated at 4°C for 18 hours. The primary antibodies used were: Anti-acetyl-Histone H4 (Lys5, 8, 12, 16) (Millipore 06-866); Anti-trimethyl-Histone H3 (Lys4) (Millipore 07-473); Anti-trimethyl-Histone H3 (Lys27) (Millipore 07-449), IgG (Santa Cruz SC-2027). 100 μl of protein A-agarose slurry was added to each sample and incubated for one hour at 4°C with rotation. Chromatin and agarose beads were then transferred to Spin-X tubes (Corning, Tewksbury MA-#8160) for washing. Each wash (600 μl) was preformed for five minutes at RT on a rotator, followed by a one-minute spin at 300 x g. Wash conditions were as follows: 1 x Low Salt Wash (20mM Tris pH 8.1, 150mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS, protease inhibitors), 1 x High Salt Wash (20mM Tris pH 8.1, 500mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS, protease inhibitors), 1 x LiCl Wash (10mM Tris pH 8.1, 250mM LiCl, 1% NP-40, 1% Na Deoxycholate, 1mM EDTA, protease inhibitor), 2 x TE wash (10mM Tris pH 8.0, 1mM EDTA, protease inhibitor). Chromatin elution was achieved by incubating the agarose beads in 200 μl Elution buffer (100mM $NaHCO_3$, 1% SDS) for 15 minutes at RT with rotation. Supernatant was collected and NaCl added to a final concentration of 200mM. Samples were then incubated at 65°C for four hours to

reverse the cross-links. Twenty micrograms of Proteinase-K was added and the buffer concentration increased to 50mM Tris and 10mM EDTA. Samples were incubated for one hour at 45°C. DNA was isolated using the ChIP DNA Clean and Concentrator Kit (Zymo Research D5205) and eluted in 50 μl TE.

**qPCR Analysis of ChIP DNA.** Prior to Illumina sequencing, each ChIP-DNA sample was first analyzed to assess quality via qPCR using an Applied Biosystems 7900HT Fast Real-Time PCR System. Each reaction was performed in triplicate and contained one microliter ChIP DNA, five microliters SyberGreen Master Mix (Applied Biosystems), and 300 nM forward and reverse primer. Primers for the *Fshr* promoter are FHSR Prom F 5'-TTTACTTGCCTGGAAGCGACTAA-3' AND FSHR Prom R 5'-GCTTGGAGAACGGGCAAA-3' and for the *Nrxn1* promoter are Nrxn1 Prom F 5'-GGGAGGAATCTGATCACTGTACTGT-3' AND Nrnx1 Prom R 5'-CATGATCTTCCCAATTGTCCAA-3'. The standard cycle was used, without any changes to the default parameters. A dissociation curve was run following the final cycle to determine if contamination was present. The SDS Software package 2.4 (Applied Biosystems) was used to calculate the cycle threshold (Ct) value for each primer set, which were adjusted such that the threshold was set at the initial phase of linear PCR amplification. Data was exported to a Microsoft Excel spreadsheet and analyzed as a percent of the input sample [100*2^(Adjusted input - Ct sample)]. Adjusted input was calculated based on the Ct value of the 1% input sample. qPCR was also used to verify regions identified by next generation sequencing.

**ChIP-DNA Sequencing.** ChIP DNA samples from two independent samples were pooled, analyzed via qPCR, dried and shipped to Cofactor Genomics LLC (St. Louis, MO) for library preparation and sequencing. Samples were prepared following Illumina library preparation protocols by Cofactor and sequenced on the Illumina GAIIx. The NEBNext® ChIP-Seq Library Prep Reagent Set for Illumina® (New England Biolabs), was used by Cofactor to prepare all samples for sequencing. Briefly, 10 ng ChIP DNA was treated with klenow and T4 DNA Polymerase to repair fragmented ends, cleaned and ligated to the NEBNext Adaptor sequence. The DNA was then cleaned and PCR amplified for 15 cycles using the NEBNext High-Fidelity 2X PCR Master Mix and Universal PCR Primers per the manufacturers protocol. The resultant library was cleaned using AMPure XP Beads, diluted and analyzed on the Agilent Bioanalyzer. Twenty picomoles of Library DNA were prepared for sequencing using the Illumina Cluster station. This ligated a llumina-supplied adapter sequence to one end of the DNA fragment, amplified the DNA clusters and hybridized the DNA to the flow-cell. The flow-cell was then loaded onto the Genome Analyzer IIx. Sequencing data output was in FASTQ files.

**Processing and Analysis of Sequencing Data.** ChIP-Seq data was analyzed through the K-INBRE Bioinformatics Core, University of Kansas, by Aaron Smalter Hall, Ph.D. FASTQ files representing eight lanes of single-end Illumina ChIP-Seq sequence tags produced by Cofactor Genomics were analyzed. FastQC (Babraham Bioinformatics-http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to test the quality of the sequence data for each sample prior to alignment.

FastQC is a quality control application for high throughput sequence data that provides an interactive application to review the results of several different QC checks relating to the quality of sequence tags. The Trim Galore! (Babraham Bioinformatics -http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) tool was used to trim low-quality base pairs, the Illumina primer and known adapter sequences using the following parameters: Quality Phred score cutoff: 51, Quality encoding type selected: ASCII+64, Adapter sequence: 'GATCGGAAGAGCGGTTCAGCAGGAATGCCG', Maximum trimming error rate: 0.1 (default), Minimum required adapter overlap (stringency): 1 bp and Minimum required sequence length before a sequence gets removed: 20 bp. The following command line was used: -f fastq -e 0.1 -q 51 -O 1 -a GATCGGAAGAGCGGTTCAGCAGGAATGCCG 'file name'.fq. Removal of duplicate sequence tag was performed using the "rmdup" module in the Samtools package (Sourceforge.net- http://samtools.sourceforge.net/). The BWA alignment tool (Sourceforge.net- http://bio-bwa.sourceforge.net/) was used to align the resulting files to the latest build of the *Rattus norvegicus* genome (rn5, May 2012) using the default parameters. Alignment was performed against two base library builds – a 'Soft Masked' and a 'Hard Masked'. The soft masked file contains the entire rn5 genome with repeat region denoted by lower case font while the hard masked genome masks all repeat regions to N's. These will be referred to as Unmasked and Masked respectively. Visualization of the mapped sequence tags was achieved using the UCSC Genome Browser using the BED, BEDGraph and BigWig file formats.

**UCSC Genome Browser Custom Track Settings.** To visualize the aligned sequence tags, bigwig files were imported into the UCSC Genome Browser using the 'add custom track' feature. The bigWig format is for display of dense, continuous data that will be displayed in the Genome Browser as a graph. Each track was imported based on the alignment of the sequence tag, i.e. positive or negative strand alignment. The following import command was used: track type=bigWig name="file name" description="" bigDataUrl=http://[ftp location of file]. Visualization settings were set as follows- Display mode: Full, Type of graph: bar, Track height: 40 pixels, Vertical viewing range: min-1 max-30, Data view scaling: use vertical viewing range setting, Always include zero: off, Transform function: Transform data point by: None, Window function: maximum, Smoothing window: off, Dray y indicator lines: at y=0.0: off at y=0 off.

## Results/Discussion

**qPCR Analysis of ChIP DNA.** ChIP-DNA was prepared from Sertoli/Germ and myoid cell chromatin immunoprecipitated with antibodies to H4ac, H3K4me3 and H3K27me3. To ensure our ChIP-DNA was of good quality for sequencing, the DNA was first assessed by qPCR to determine the histone modifications associated with the *Fshr* and the *Nrxn1* promoter. These two promoter regions were selected for analysis because of their known expression patterns. *Fshr* is expressed only in Sertoli cells and *Nrxn1* is not expressed in either cell type. The *Fshr* promoter was associated with H4ac enrichment in both cell types, with slightly higher levels observed in myoid cells (Figure 2-1). Within Sertoli cells, enrichment of H3K4me3 was highest at the *Fshr* promoter; however, H3K27me3 enrichment was highest in myoid cells. The *Nrxn1* promoter was not associated with enrichment of any histone modifications tested in either cell type. We didn't expect the *Nrxn1* promoter to have much enrichment for any modification tested as this region is likely heterochromatic in both cell types. For *Fshr* we did expect the inverse relationship of enrichment between the two cell types as H3K4me3 associates with active promoters (*Fshr*) and H3K27me3 associates with inactive promoters (*Nrxn1*). We concluded from this analysis that our ChIP-DNA was of good quality and continued with DNA sequencing.

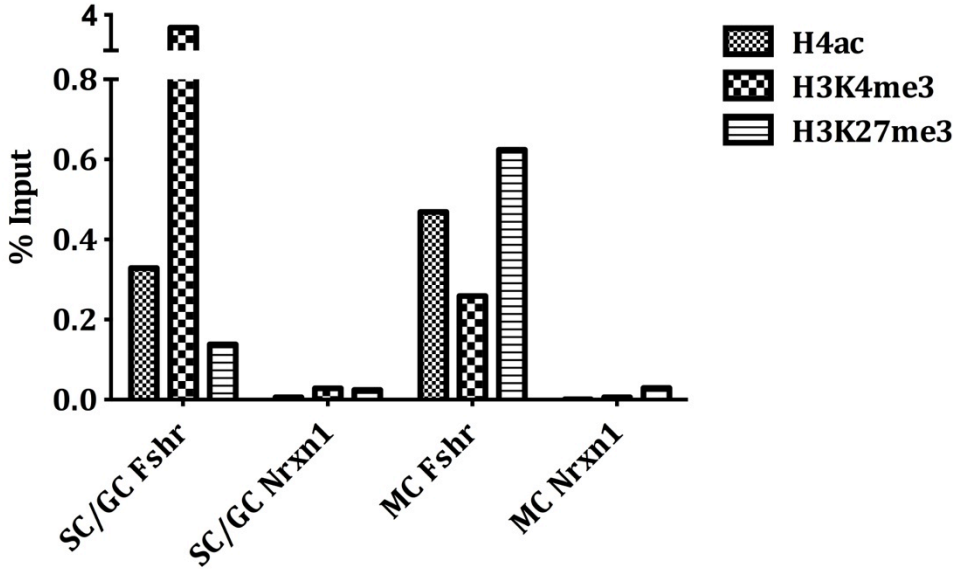



**Figure 2-1. Analysis of *Fshr* and *Nrxn1* promoters in ChIP-DNA samples.** Promoter regions were evaluated by qPCR using Sertoli/Germ cell (SC/GC) and myoid cell (MC) immunoprecipitated chromatin identifying Histone 4 acetylation, Histone 3 Lysine 4 tri-methylation and Histone 3 Lysine 27 tri-methylation.

**Illumina Sequencing of ChIP DNA.** ChIP-DNA was sent to Cofactor Genomics (St. Louis, MO) for library preparation and sequencing on the Genome Analyzer IIx (Illumina Inc., San Diego, CA). The total number of sequencing tags generated for each ChIP sample is listed in Table 2-1. The Illumina GAIIx sequencer is capable of generating 12-30 million sequence tags per lane and all of our generated tag counts fall within this range.

The Input sample for the Sertoli cells produced roughly 3 million fewer sequence tags than the input sample for myoid cells. As each Input sample was generated using the same amount of chromatin for each cell type, we expected each to produce a similar amount of sequence tags. Although we typically think of micrococcal nuclease digest as equally fragmenting chromatin, this is not necessarily true. Micrococcal nuclease not only cleaves at nucleosomes, but also will cleave naked DNA as well as sequences rich in AT/TA dinucleotides (Becker, Yau et al. 2013), which can lead to an over and/or underrepresentation of certain genomic sequences. Micrococcal nuclease also tends to digest euchromatin better than heterochromatin (Yin, Sweeney et al. 2011). Since Sertoli cells are fully differentiated and non-dividing, they likely contain more heterochromatic regions than the myoid cells. This could lead to a bias in our fragmentation and thus be the cause of fewer sequence tags being identified in the Sertoli cell input sample.

The total number of sequence tags generated for the Sertoli cell samples H4ac and H3K4me3 are several million bases lower than in the Myoid cell sample. As noted above, the Sertoli cells likely contain more heterochromatin due to their differentiation state when compared to myoid cells. This could explain why fewer

sequence tags were generated in the Sertoli cell sample for H4ac and H3K4me3, as these marks are commonly associated with euchromatic active regions.

For our studies, we used H3K27me3 as a mark to identify repressed regions in the genome. The sequencing tags generated for the H3K27me3 samples were similar across both cell types. A study by Young *et. al.* identified similar amounts of H3K27me3 enrichment across two cell types (mouse Embryonic Stem Cells and $G_1$ME cells) yet noted that the locations of enrichment differed within each cell type (Young, Willson et al. 2011). This could indicate that Sertoli/Germ cells and myoid cells share a similar number of silenced or repressed genes.

**Table 2-1. Total sequence tags generated from Illumina sequencing.**

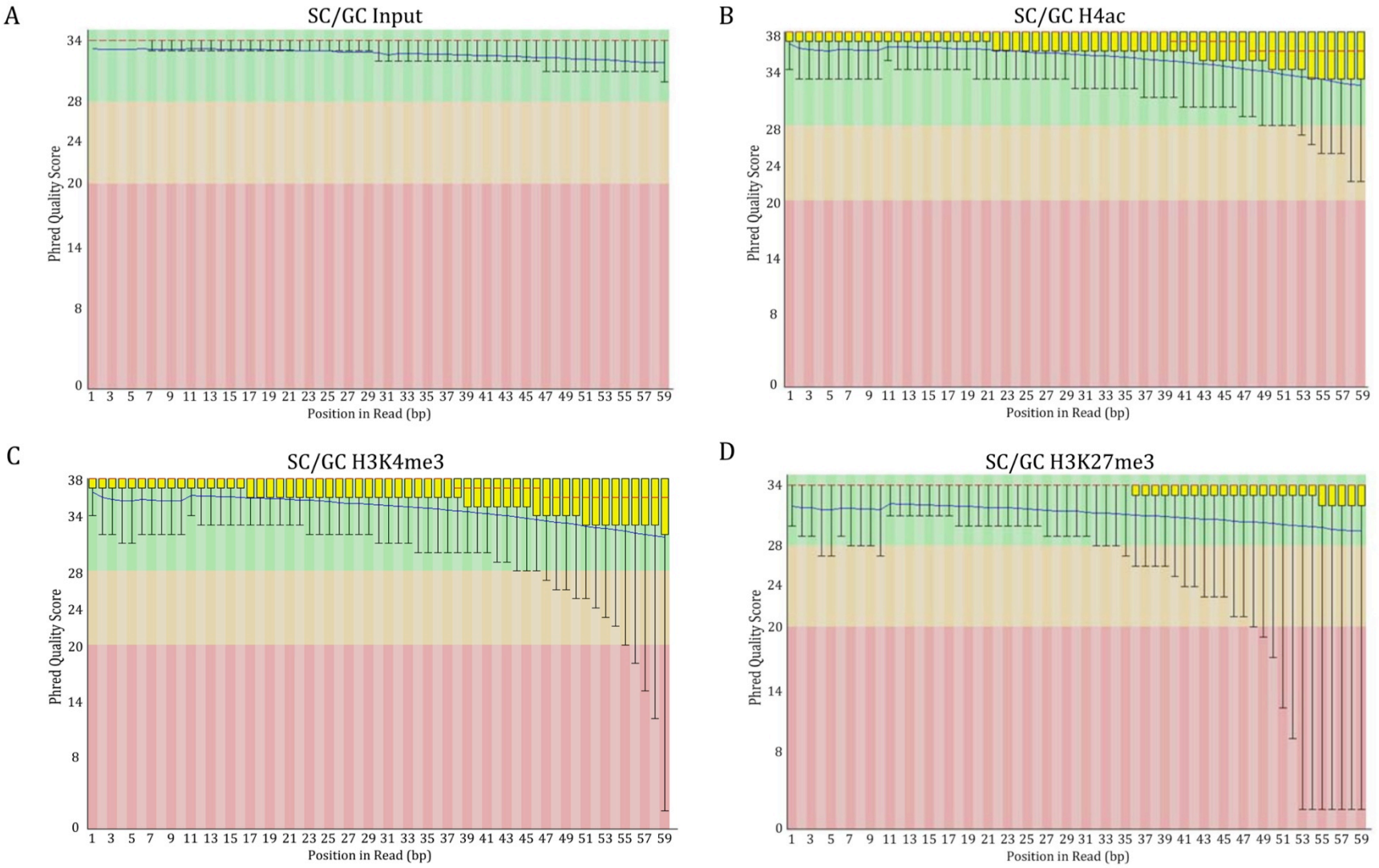
| | ChIP Sample | Illumina Raw Seq Tags |
|---|---|---|
| **Sertoli/ Germ Cells** | Input | 21,384,126 |
| | H4ac | 15,285,827 |
| | H3K4me3 | 14,888,624 |
| | H3K27me3 | 19,192,554 |
| **Myoid Cells** | Input | 24,406,389 |
| | H4ac | 20,022,701 |
| | H3K4me3 | 18,352,272 |
| | H3K27me3 | 19,072,572 |

**Quality Control Analysis to Assess Sequence Tag Quality.** Before aligning the sequence data for each sample to the reference genome, the sequence was analyzed using FastQC, a software program that tests the quality of raw sequence data. The program analyzes multiple parameters assess problems that need corrected before alignment a reference genome. Each ChIP-Seq sample was analyzed independently, for eight different parameters, each of which is discussed below followed by the graphic data for each report.

*Per Base Sequence Quality*

The per base sequence quality report gives an overview of the quality score across all bases at each position in the sequence tag, across all sequence tags in the sample. The quality score is determined using the Phred score assigned to each base call, which is determined by the Illumina sequencer (Ewing and Green 1998; Ewing, Hillier et al. 1998). The quality score data is graphed using standard box-whisker plots. The central red line is the median value, the yellow box represents the inter-quartile range (25-75%), the upper and lower whiskers represent the 10% and 90% points and the blue line represents the mean quality. A Phred score of 20 indicates a base call accuracy of 99%, a score of 30 indicates accuracy of 99.9% and a score of 40 indicates accuracy of 99.99% at each base (Ewing, Hillier et al. 1998). A quartile Phred value of less than 10, or a median Phred value of less that 25 at any base will cause a warning, and a quartile Phred value of less than 5 or a median Phred value of less than 20 at any base will cause a failure. Each sequenced ChIP sample passed this quality report indicating that each base called during sequencing was of good quality (Figure 2-2 A-H).

Although the SC H3K4me3 and SC H3K27me3 samples passed this report, it is important to note that the 3' bases of these sequences contain base calls that are suboptimal (Figure 2-2 C and D). The SC H3K4me3 sample disparity is refined to the distal four bases but the SC H3K27me3 sample disparity includes the distal 11 bases. It is difficult to determine from these graphs how many of sequence tags within each sample contain poor Phred scores at the 3' end. However, should these samples align poorly, trimming of the 3' bases may increase the number of sequence tags that align to the reference genome.


A
SC/GC Input
B
SC/GC H4ac
C
SC/GC H3K4me3
D
SC/GC H3K27me3
Phred Quality Score
Position in Read (bp)

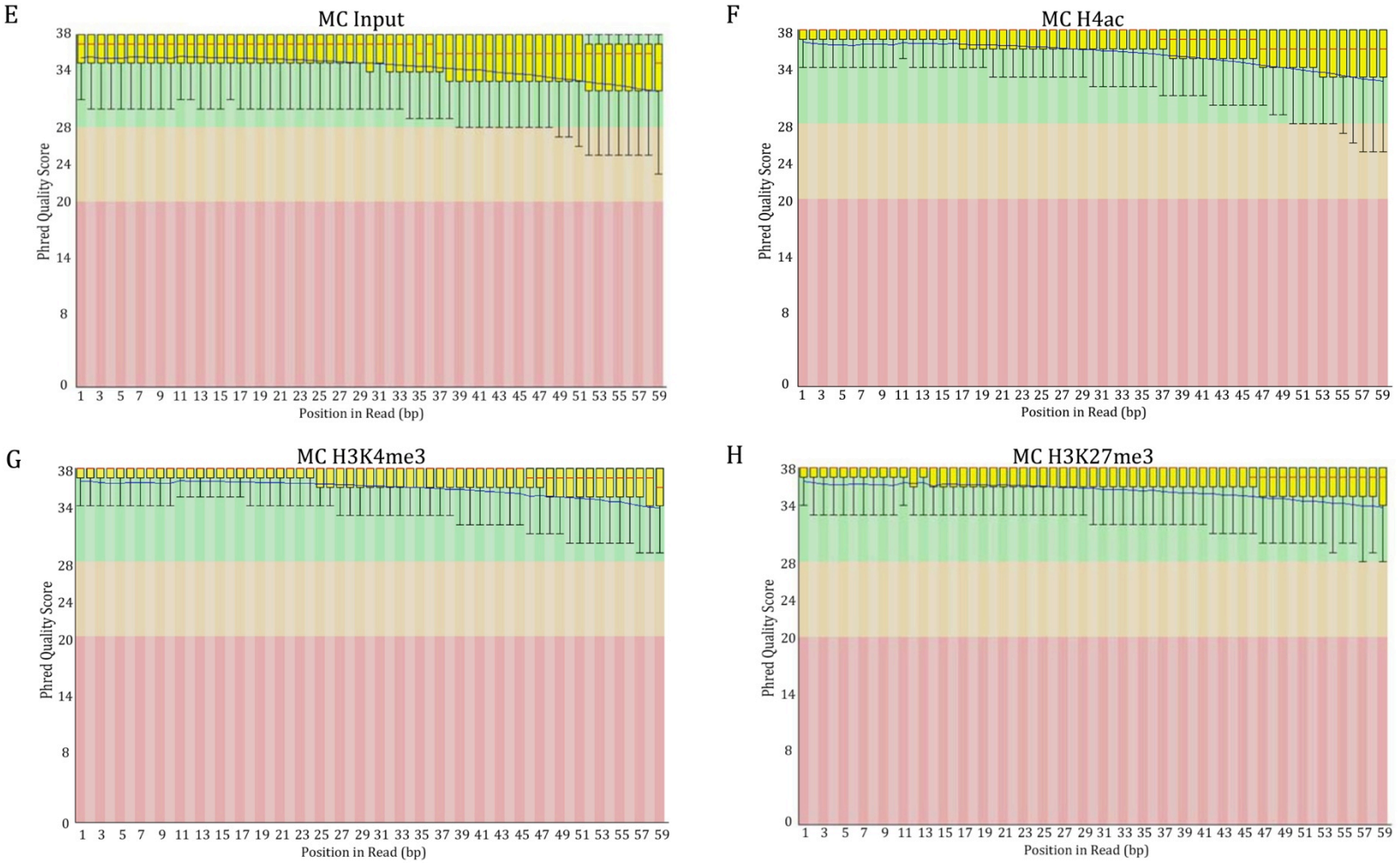


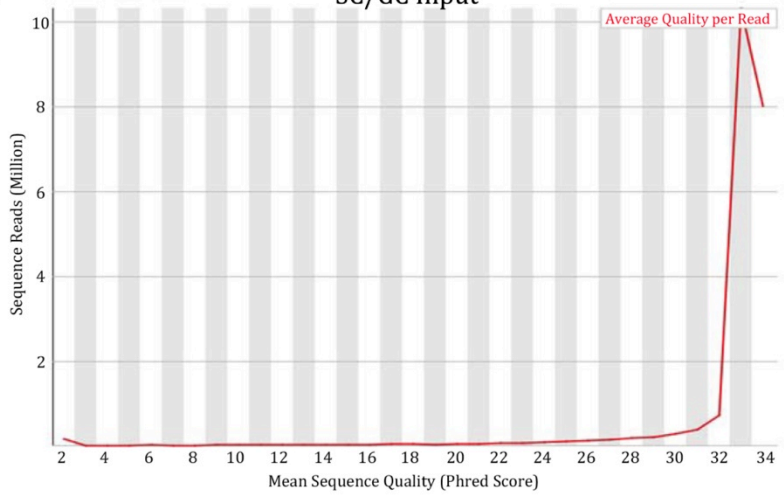

**Figure 2-2. Per base sequence quality report generated from FastQC analysis.** Quality scores across all bases for Sertoli/germ (A-D) and myoid cell (E-F) Input (A, E), H4ac (B, F), H3K4me3 (C, G) and H3K27me3 (D, H). The data are graphed using standard box-whisker plots. The central red line is the median value, the yellow box represents the inter-quartile range (25-75%), the upper and lower whiskers represent the 10% and 90% points and the blue line represents the mean quality.

*Per Sequence Quality*

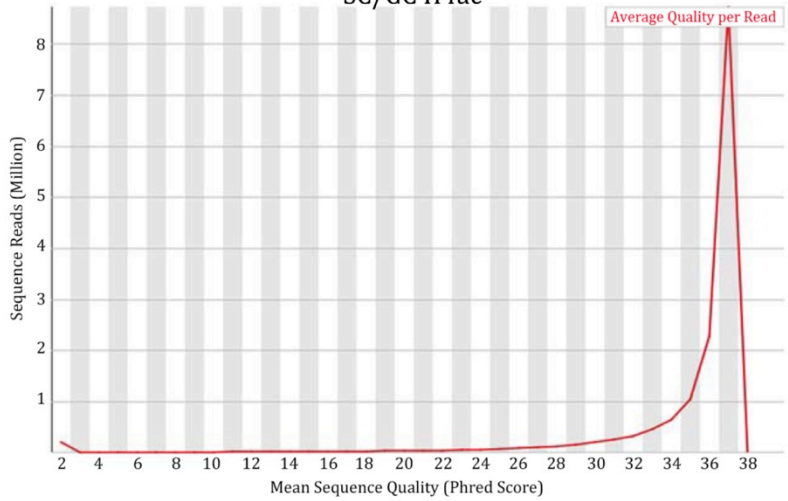
The per sequence quality report uses the average Phred quality scores across a sequence tag to determine if a subset of your sequence tags have overall low score values. Graphic reports identify the total number of sequence tags (y-axis) and the average Phred score across each sequence tag (x-axis). The average Phred score for each sequence tag was calculate by averaging the Phred scores at each base call across the sequence tag length (59 bases). Samples failing this control test indicate there was a problem during the sequencing run, usually attributed to poor starting DNA quality. If the most frequently observed Phred score is below 27 (=0.2% error rate), then a warning will be issued. If the most frequently observed Phred score is below 20 (=1% error rate), then a failure will be issued. Each sequenced ChIP sample passed this quality report indicating that each sequence is of good quality (Figure 2-3 A-H).

Two samples did display a slight increase associated with a very low mean sequence quality (Mean sequence quality of 2 on the x-axis): SC H3K4me3 and SC H3K27me3 (Figure 2-3 C and D). Roughly 300,000 and 800,000 sequence tags respectively are of poor sequence quality and may be difficult to align to a reference genome. The poor quality is likely due to the poor quality scores associated with the 3' bases of these sequences as outlined above in the per base sequence quality report. Again, this data indicates that the 3' bases of these sequence tags may need trimmed prior to alignment to a reference genome.

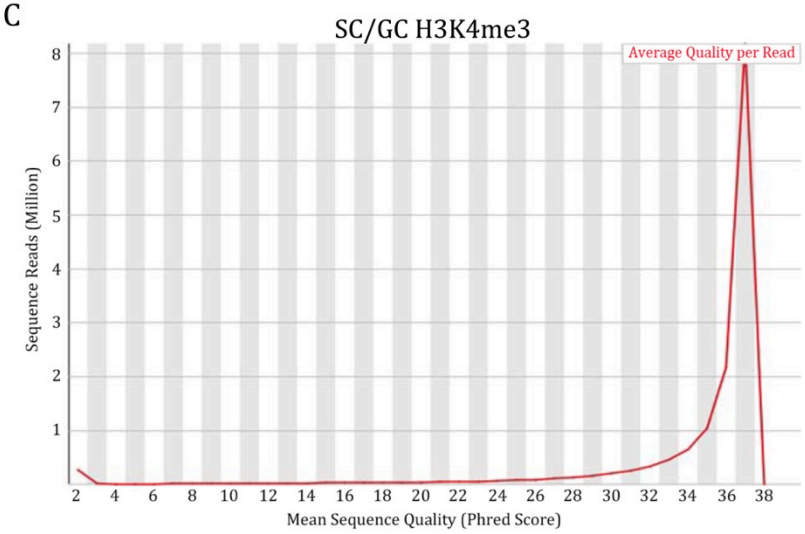
A


SC/GC Input
Average Quality per Read
Sequence Reads (Million)
Mean Sequence Quality (Phred Score)

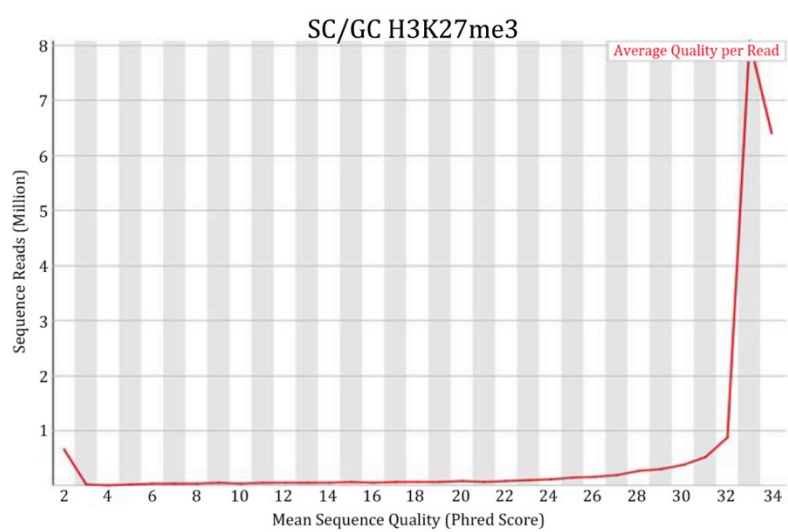
B


SC/GC H4ac
Average Quality per Read
Sequence Reads (Million)
Mean Sequence Quality (Phred Score)

C


SC/GC H3K4me3
Average Quality per Read
Sequence Reads (Million)
Mean Sequence Quality (Phred Score)

D


SC/GC H3K27me3
Average Quality per Read
Sequence Reads (Million)
Mean Sequence Quality (Phred Score)

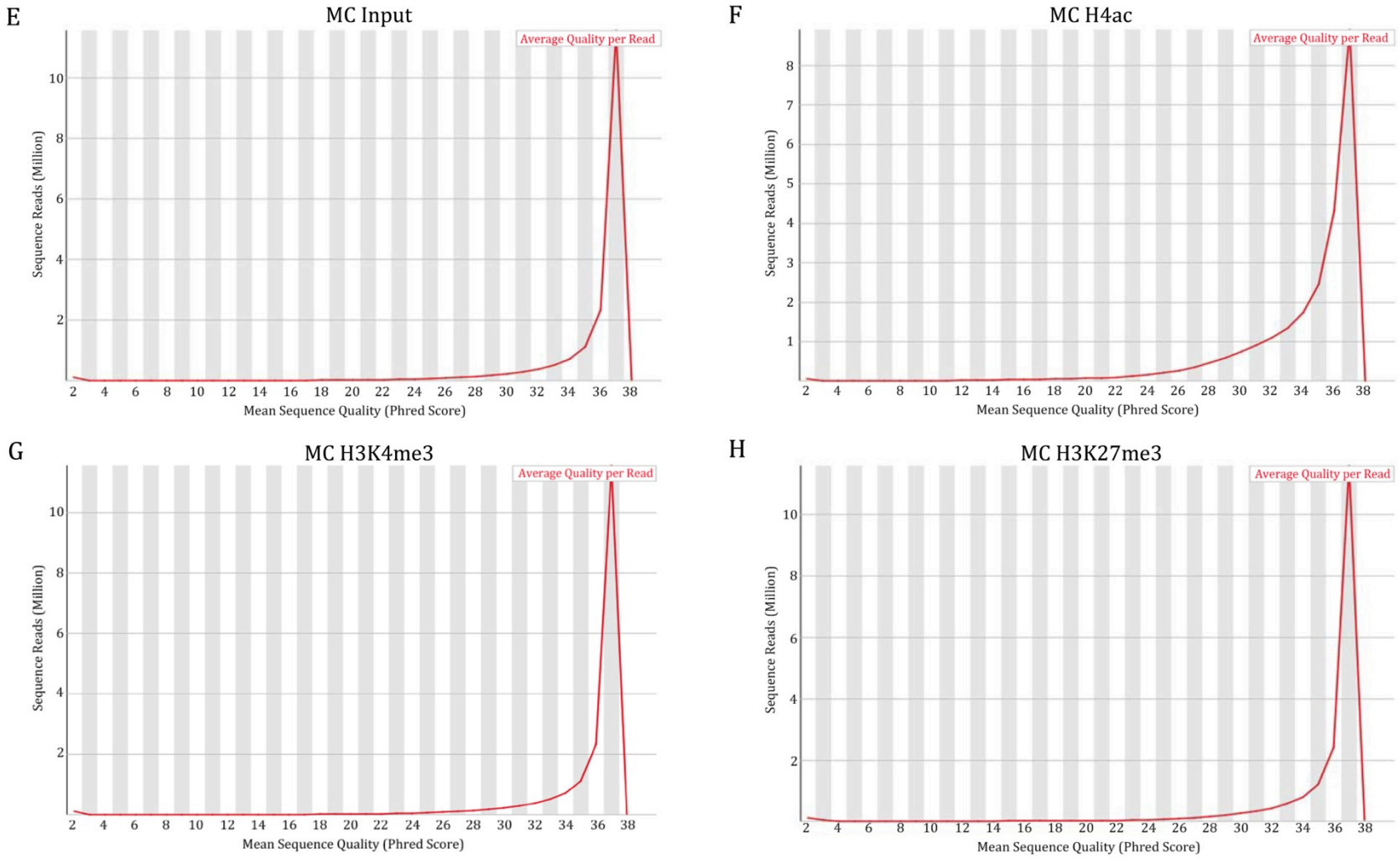



**Figure 2-3. Per sequence quality report generated from FastQC analysis.** Quality score distribution over all sequences for Sertoli/germ (A-D) and myoid cell (E-F) Input (A, E), H4ac (B, F), H3K4me3 (C, G) and H3K27me3 (D, H). The average Phred score was calculated for each sequencing read and graphed (red line).

*Per Base Sequence Content*

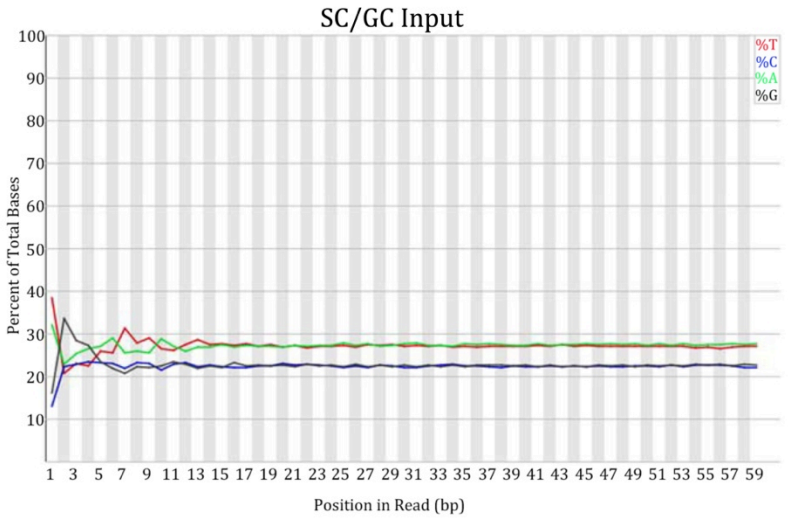
The per base sequence content was graphed as the proportion of each called base across each position of the sequence tag. The percentage of each base called (A, T, C or G) is graphed (y-axis) corresponding each base call across the sequence tag (x-axis, 59 bp). A strong bias of any base at any position indicates the presence of overrepresented sequence. Theoretically, the proportion of bases in each position across the sequence tag should be the same. A failure of this quality report indicates that either the original library was biased, or that there was an overrepresented sequence occurring in the sample. An overrepresented sequence can be identified by a strong bias associated with specific bases, and if the original library had sequence bias, the bias association will occur across all the bases within the sequence tag. Bias is usually attributed to the sequencing of linkers or adapters within the 5' or 3' bases of a sequence tag, or could occur across the sequence tag if there is an overrepresented sequence present that is likely caused by PCR duplicates.

This report will issue a warning if the difference between A and T, or C and G is greater that 10% at any base or a failure if the difference is greater than 25% at any base. Each sequenced ChIP sample either failed or demonstrated a warning for this quality report (Figure 2-4 A-H). However, upon further review, the disparity that occurs within our samples is always confined to the first five bases of the sequence tag. This indicates the presence of a similar sequence; most likely the sequencing primer or adapter sequence ligated to the sequence tags for library

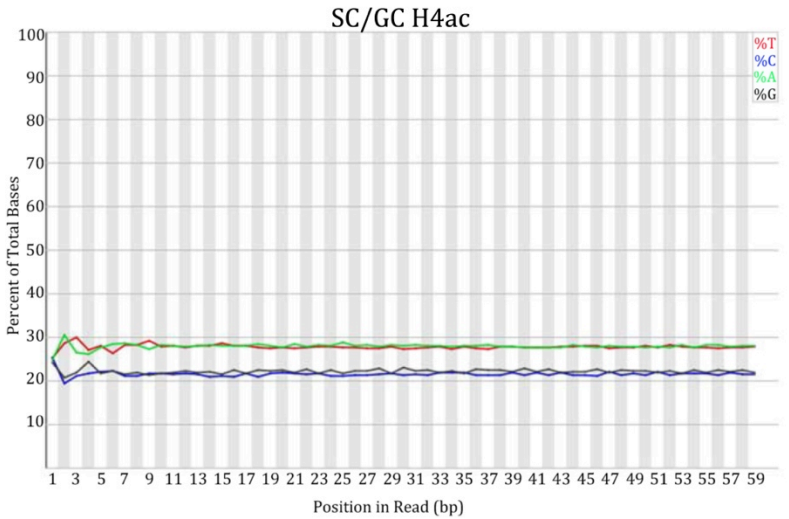
preparation. This report tells us that we need to trim the initial bases of each sequence tag before continuing with alignment to the reference genome.

When comparing the graphs beyond the first five bases, each sample is associated with equally distributed bases calls except the MC H3K27me3 sample (Figure 2-4 H). This graph identifies a disparity of ~4% between the A/T and G/C base calls across the length of the sequence tag. This indicates the presence of numerous similar sequences within the sample. These similar sequences often arise from PCR duplicates generated during the sequencing run and should be filtered out prior to genome alignment. Failure to remove duplicate sequences could lead to false identification of enriched regions when using peak finding software.

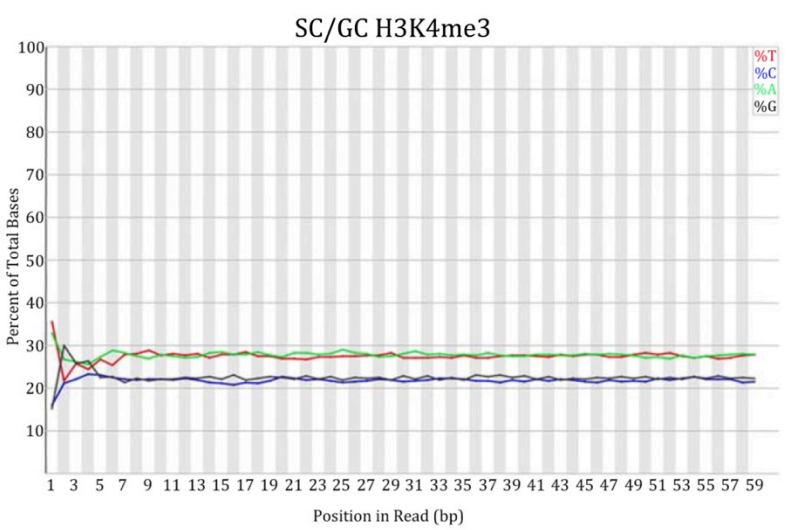
A


SC/GC Input
%T
%C
%A
%G
Percent of Total Bases
Position in Read (bp)

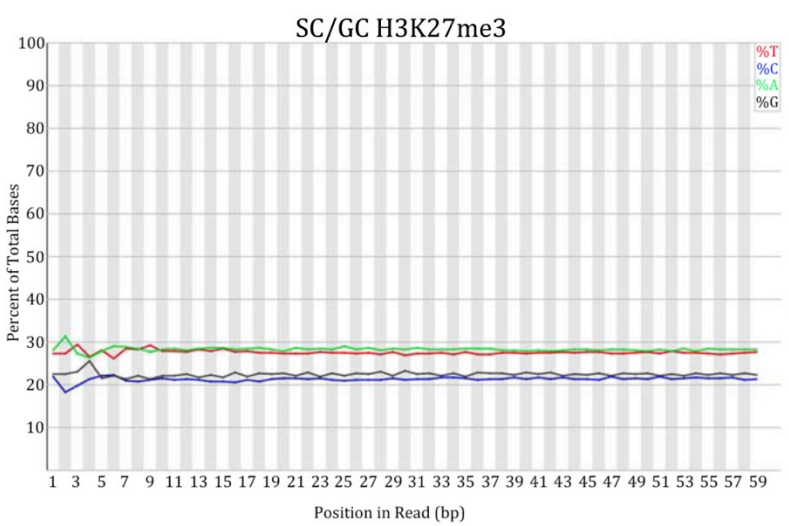
B


SC/GC H4ac
%T
%C
%A
%G
Percent of Total Bases
Position in Read (bp)

C


SC/GC H3K4me3
%T
%C
%A
%G
Percent of Total Bases
Position in Read (bp)

D


SC/GC H3K27me3
%T
%C
%A
%G
Percent of Total Bases
Position in Read (bp)

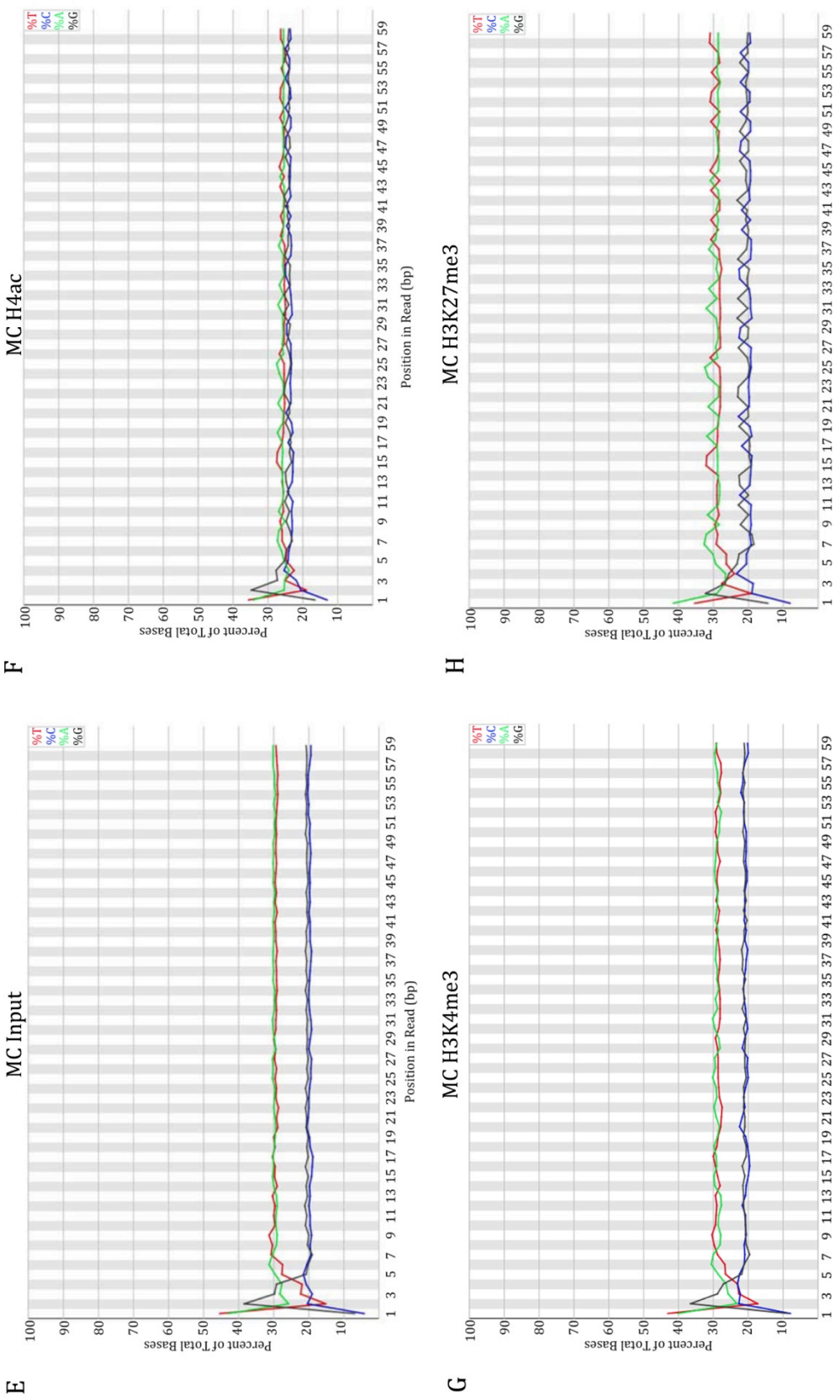


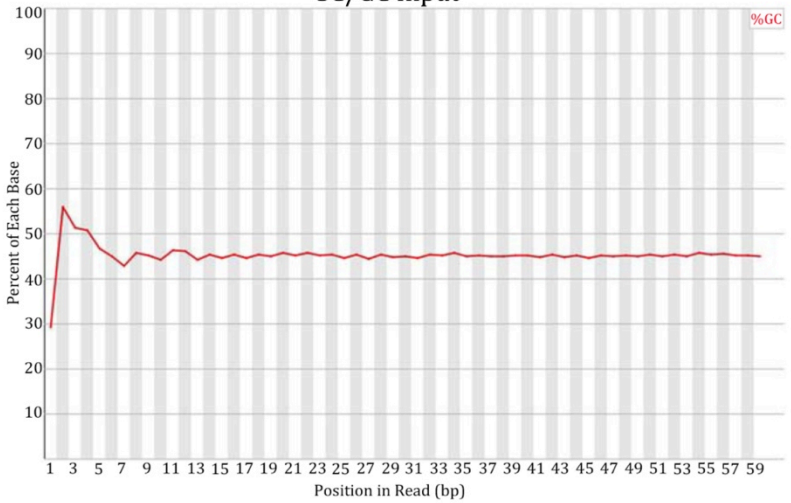

**Figure 2-4. Per base sequence content report generated from FastQC analysis.** Sequence content across all bases for Sertoli/germ (A-D) and myoid cell (E-F) Input (A, E), H4ac (B, F), H3K4me3 (C, G) and H3K27me3 (D, H).

*Per Base GC Content*

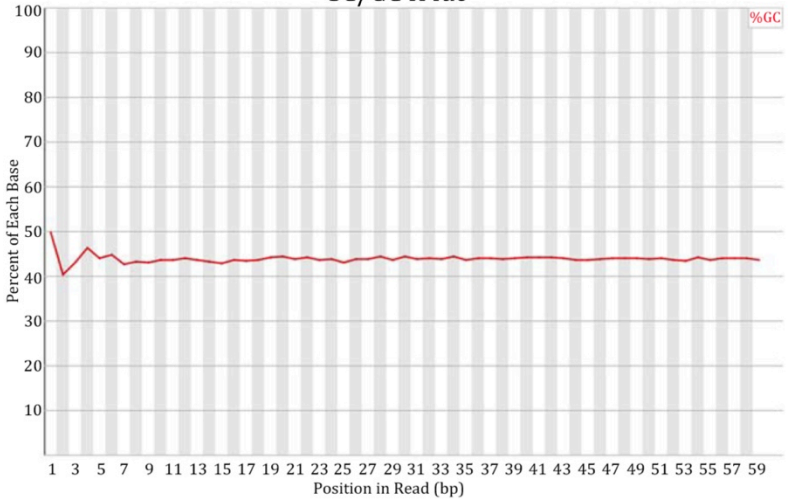
The per base GC content report was graphed as the percent of GC bases (y-axis) at each position in the sequence tag (x-axis). The GC content should reflect the GC content of the sequenced genome, which for the rat genome is ~42%, and should not change across the sequence tag (Cooper, Brudno et al. 2004; Hammond, Swanberg et al. 2011). Changes in the GC content across the sequence tag would indicate the presence of a contaminating sequence, i.e. sequencing primers or adapters, or the presence of duplicate sequence tags.

If the GC content varies more than five percent from the baseline, a warning is issued and if the GC content varies more than 10 percent a failure is issued. Only one ChIP sample passed this quality report (SC H3K27me3, Figure 2-5 D) while the others either failed or demonstrated a warning (Figure 2-5 A-H). Again, similar to the Per Base Sequence report, the disparity in the GC content that caused the warning or failure occurred within the first five to seven bases of the sequence tags; indicating the presence of a similar sequence (i.e. sequencing adapter) at the beginning of each sequence tag that needs removed prior to alignment. The MC H3K27me3 sample also indicates a variance across the length of the sequence tag implying the presence of duplicate sequences (Figure 2-5 H) that will need removed prior to alignment.

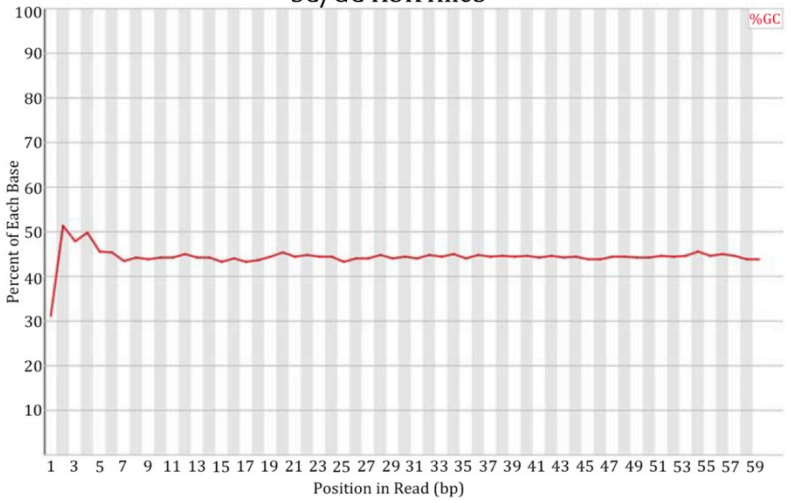
A


SC/GC Input
%GC
Percent of Each Base
Position in Read (bp)

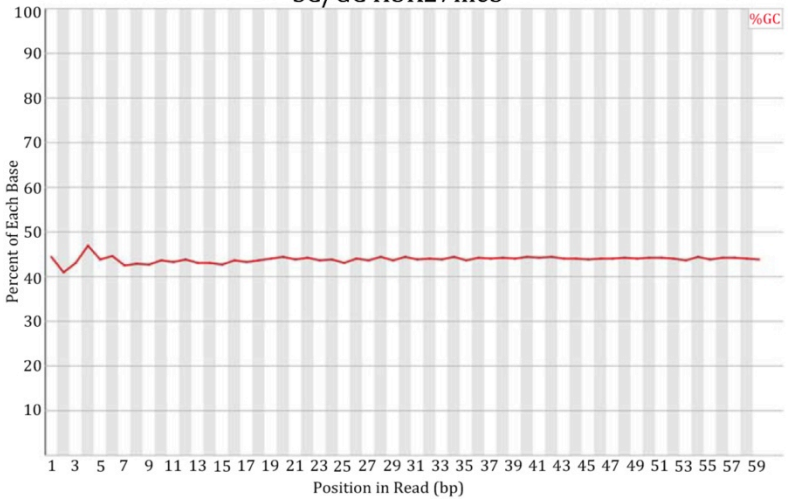
B


SC/GC H4ac
%GC
Percent of Each Base
Position in Read (bp)

C


SC/GC H3K4me3
%GC
Percent of Each Base
Position in Read (bp)

D


SC/GC H3K27me3
%GC
Percent of Each Base
Position in Read (bp)

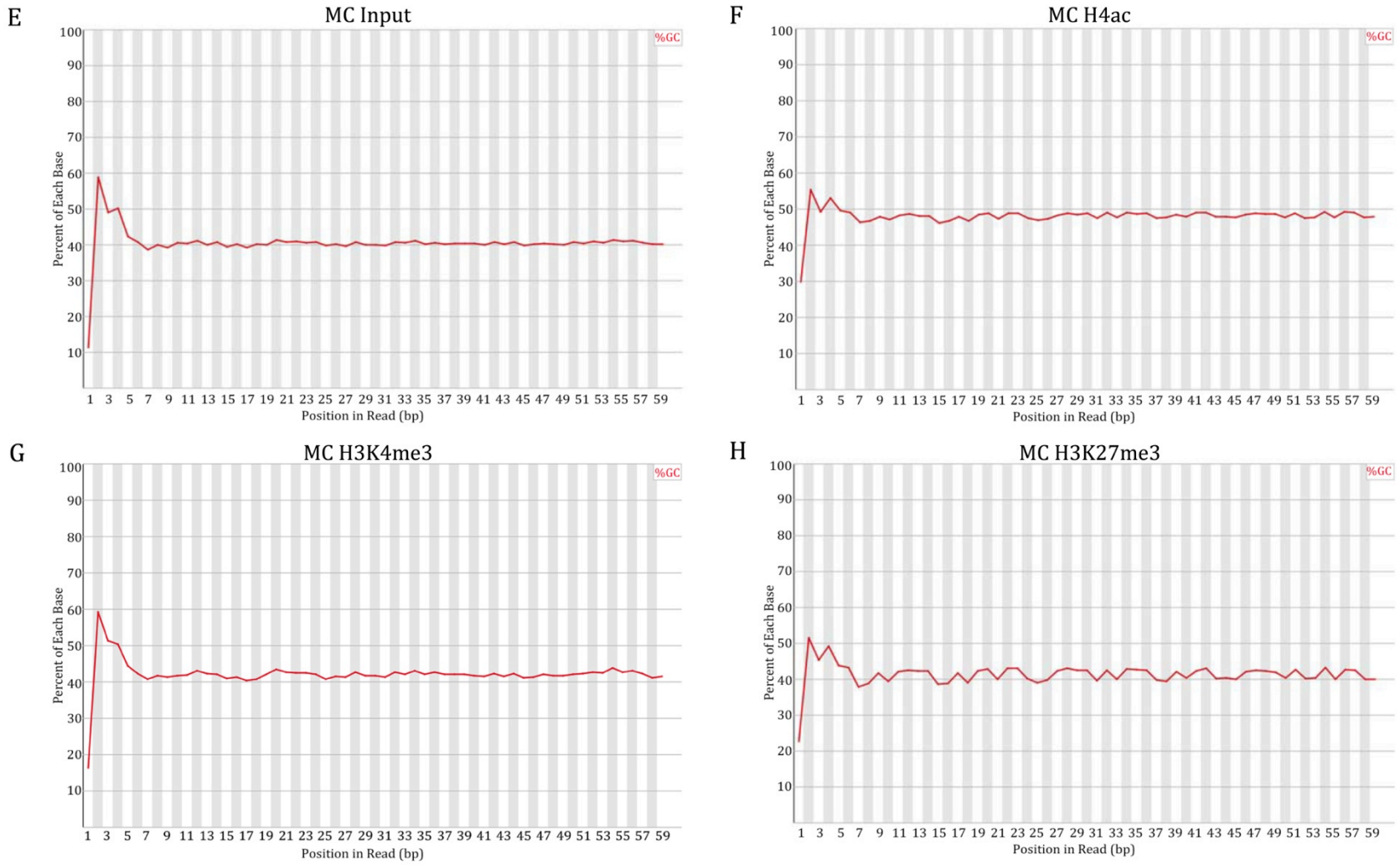
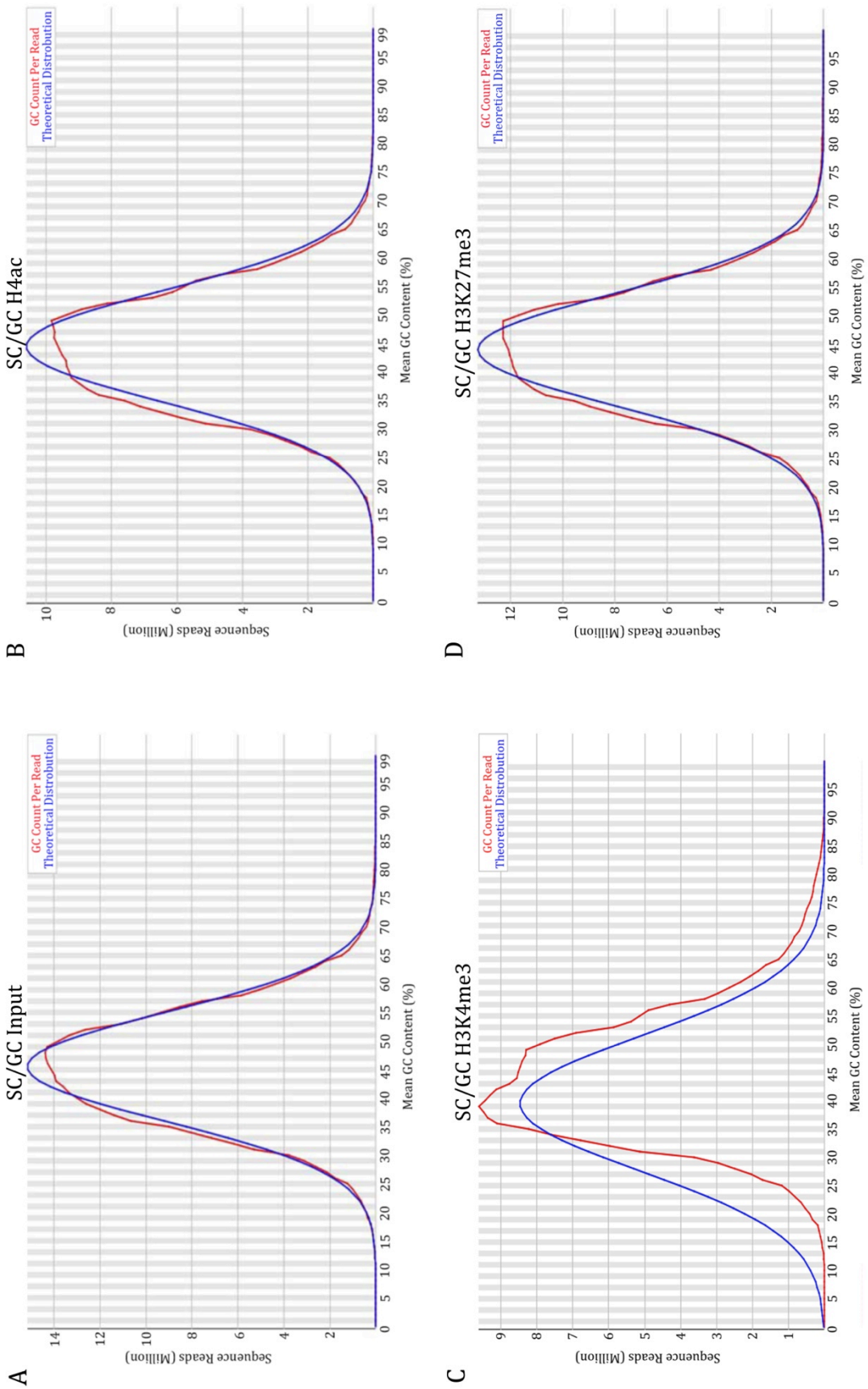



**Figure 2-5. Per base GC content report generated from FastQC analysis.** GC content across all bases for Sertoli/germ (A-D) and myoid cell (E-F) Input (A, E), H4ac (B, F), H3K4me3 (C, G) and H3K27me3 (D, H).

*Per Sequence GC Content*

The Per Sequence GC content report is graphed as a Gaussian (normal) distribution curve. It measures the GC distribution across sequence tags for each sample and compares it to a modeled theoretical distribution where the central peak corresponds to the overall theoretical GC content of the genome, which is calculated based on the sample since no reference genome is selected for this analysis. A biased sample that does not follow the normal distribution pattern suggests that the library was contaminated or there was a significantly overrepresented sequence. If the distribution deviates more than 15% a warning is issued and if the deviation is more than 30% a failure is issued. Half of our ChIP-Seq samples passed this report (SC Input, SC H4ac, SC H3K27me3 and MC Input – Figure 2-6 A, B, D and E) while the other half either generated a warning (MC H3K4me3 and MC H3K27me3 – Figure 2-6 G and H) or failed (SC H3K4me3 and MC H4ac – Figure 2-6 C and E). The distinct peaks in the GC content identified in the MC H4ac and MC H3K27me3 (Figure 2-6 F and H) indicates the presence of biased sequence/s, likely due to the presence of an overrepresented sequence, as the peak is sharp and distinct from the rest of the normal curve. The remaining samples that generated or failed are more difficult to identify the underlying cause of the failure based on the distribution of the graph. Each of these samples generated a smooth distribution curve; the curve just deviates from the theoretical, usually in the uppermost portion of the curve. This type of deviation indicates that either too many (C, E and G) or too few (A, B and D) sequence tags were associated with the theoretical average.


A
SC/GC Input
GC Count Per Read
Theoretical Distrobution
Sequence Reads (Million)
Mean GC Content (%)
B
SC/GC H4ac
GC Count Per Read
Theoretical Distrobution
Sequence Reads (Million)
Mean GC Content (%)
C
SC/GC H3K4me3
GC Count Per Read
Theoretical Distrobution
Sequence Reads (Million)
Mean GC Content (%)
D
SC/GC H3K27me3
GC Count Per Read
Theoretical Distrobution
Sequence Reads (Million)
Mean GC Content (%)

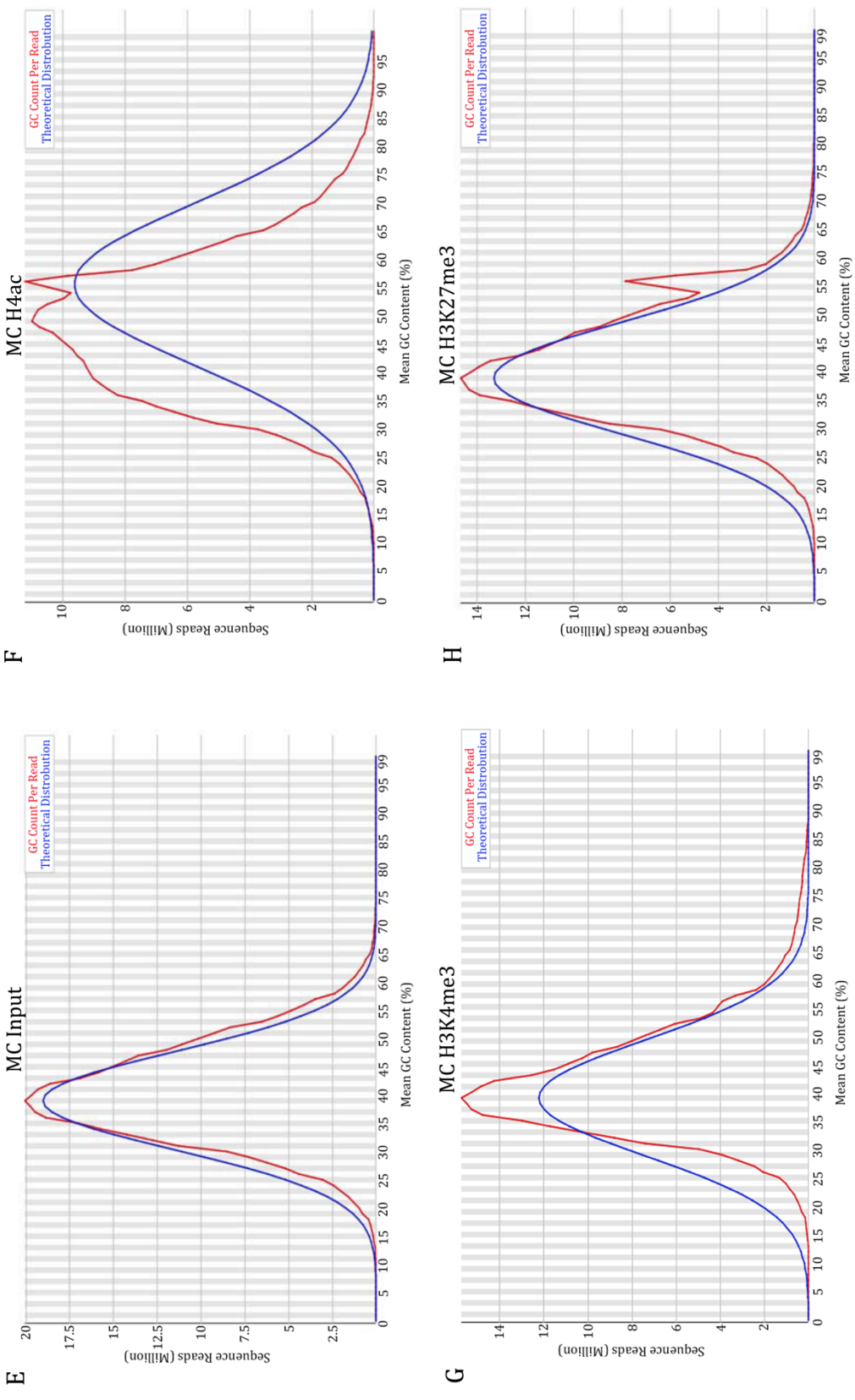
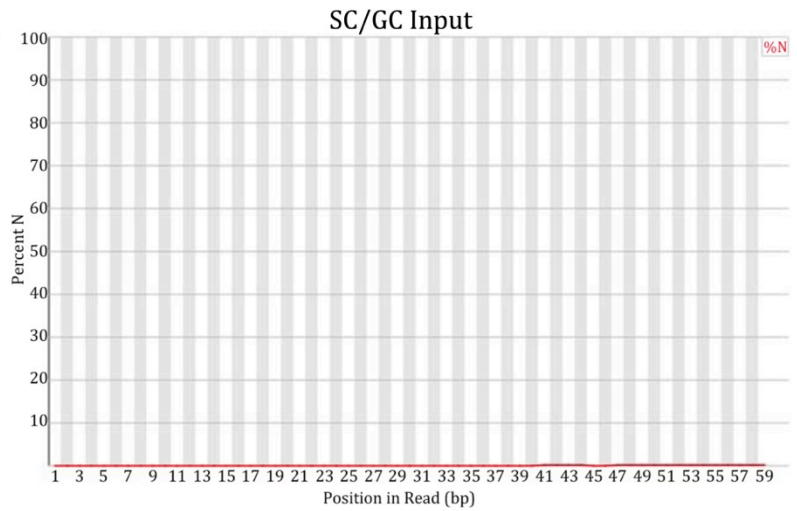



**Figure 2-6. Per sequence GC content report generated from FastQC analysis.** GC content across each sequence for Sertoli/germ (A-D) and myoid cell (E-F) Input (A, E), H4ac (B, F), H3K4me3 (C, G) and H3K27me3 (D, H).

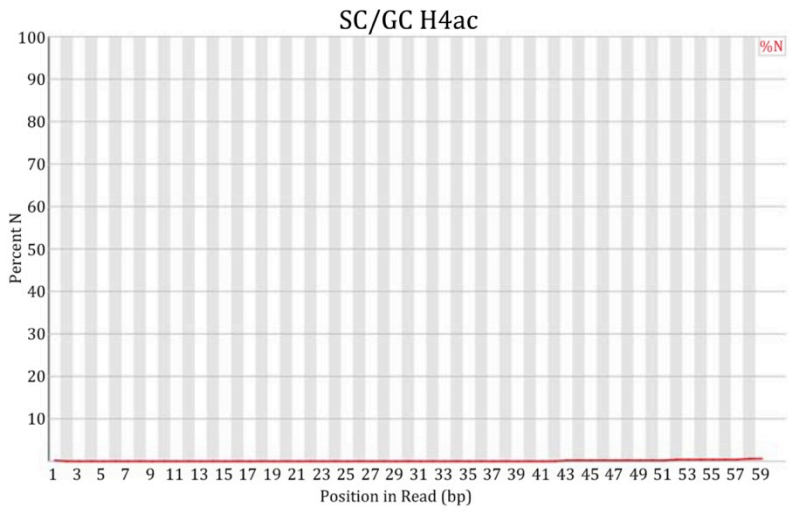
*Per Base N Content*

The Per Base N Content was graphed as the percent of bases that could not be called across each position in the sequence tag. This report summarizes the number of bases that could not be called within the sequence tag by the sequencer. A warning was issued if the percent of uncalled bases is greater than five percent at any sequence position, and a failure was issued if the N content is greater than 20% a any sequence position. Every ChIP-Seq sample passed this quality control test (Figure 2-7 A-H) indicating that there were very few bases of such poor quality that a base call was not possible.

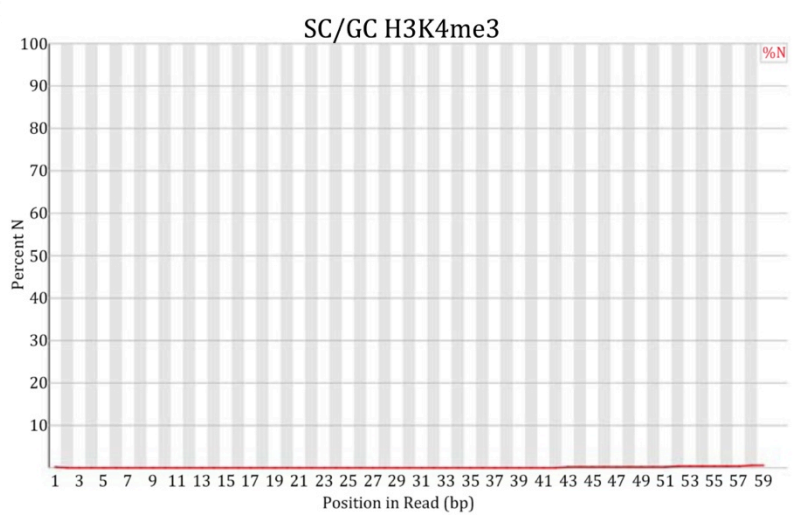
A


SC/GC Input
%N
Percent N
100
90
80
70
60
50
40
30
20
10
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Position in Read (bp)

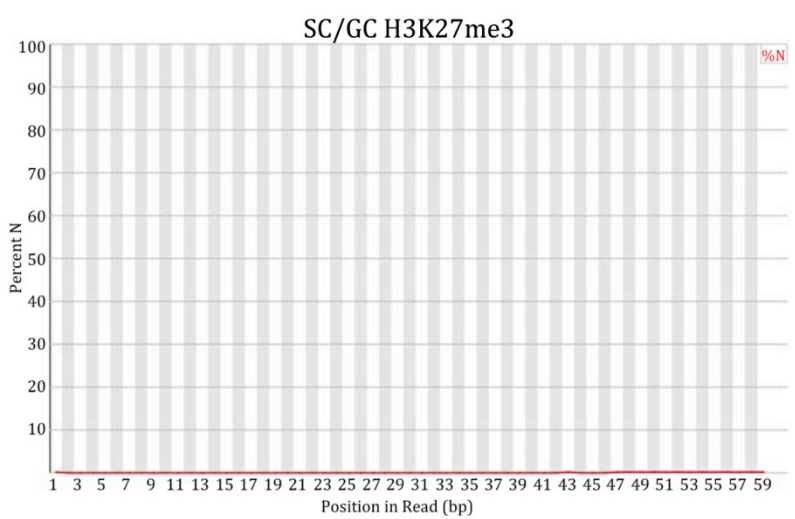
B


SC/GC H4ac
%N
Percent N
100
90
80
70
60
50
40
30
20
10
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Position in Read (bp)

C


SC/GC H3K4me3
%N
Percent N
100
90
80
70
60
50
40
30
20
10
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Position in Read (bp)

D


SC/GC H3K27me3
%N
Percent N
100
90
80
70
60
50
40
30
20
10
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Position in Read (bp)

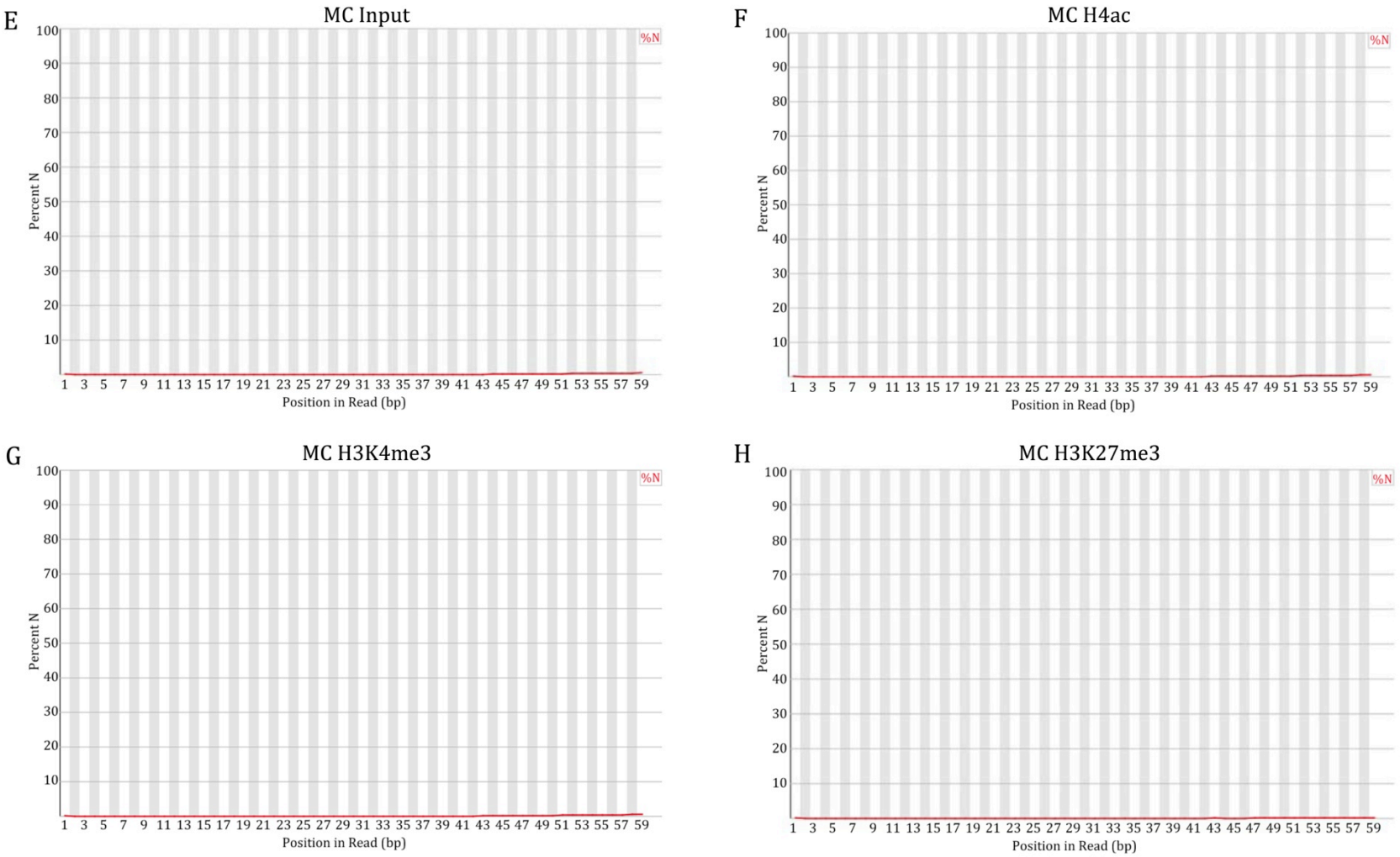
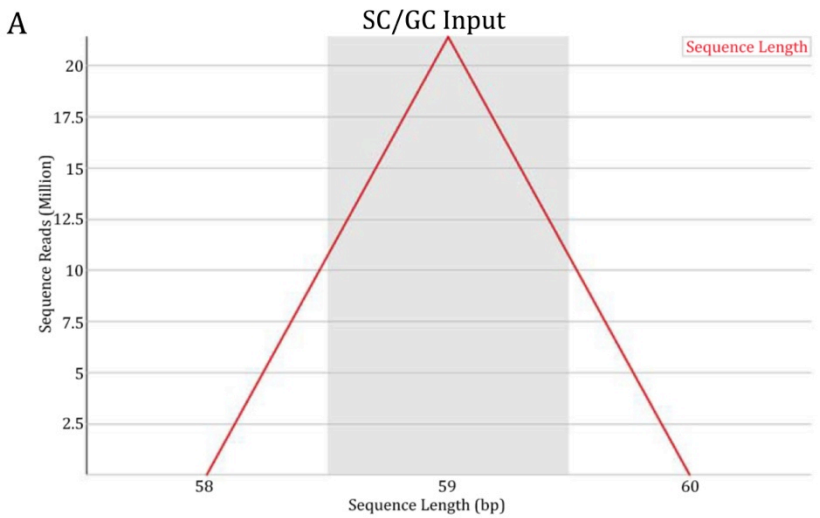
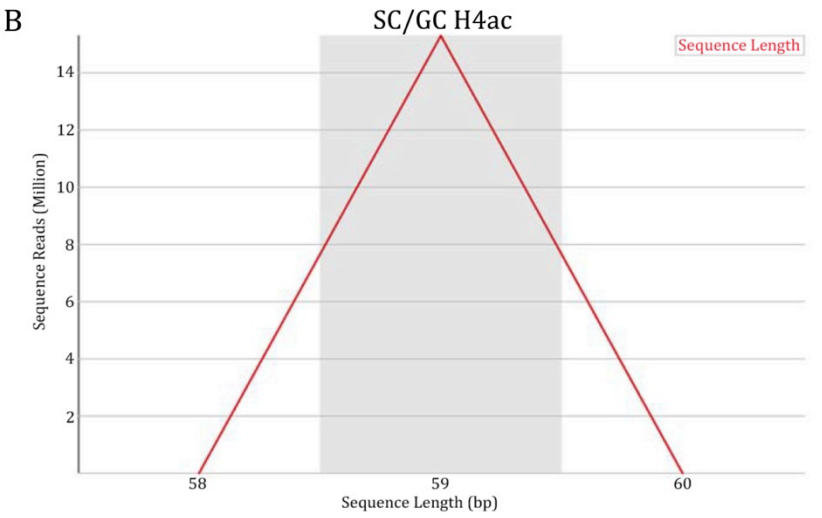


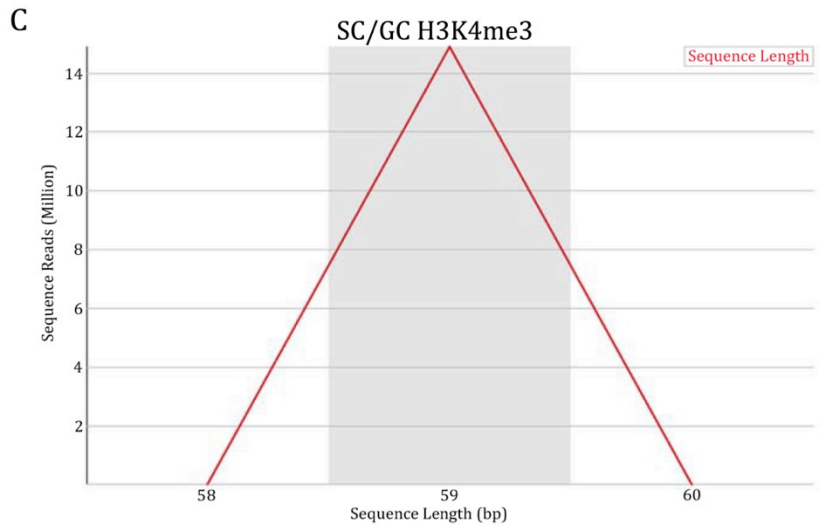

**Figure 2-7. Per base N content report generated from FastQC analysis.** N content across all bases for Sertoli/germ (A-D) and myoid cell (E-F) Input (A, E), H4ac (B, F), H3K4me3 (C, G) and H3K27me3 (D, H).

*Sequence Length Distribution*

The Sequence Length Distribution report is graphed as the number of sequence tags (y-axis) versus the length of the sequence tag (x-axis). This figure represents the length of each sequence tag represented in each ChIP sample library. As sequencing was performed on the Illumina Genome Analyzer IIx, each sequence tag should be 59 base pairs long. Every ChIP-Seq sample passed this quality control test indicating each sequence tag was the appropriate length. This tells us that there was not an issue with the sequencing run (i.e. early termination or mispriming) and all samples were adequately sequenced (Figure 2-8 A-H).


A
SC/GC Input
Sequence Length
Sequence Reads (Million)
20
17.5
15
12.5
10
7.5
5
2.5
58
59
60
Sequence Length (bp)


B
SC/GC H4ac
Sequence Length
Sequence Reads (Million)
14
12
10
8
6
4
2
58
59
60
Sequence Length (bp)


C
SC/GC H3K4me3
Sequence Length
Sequence Reads (Million)
14
12
10
8
6
4
2
58
59
60
Sequence Length (bp)


D
SC/GC H3K27me3
Sequence Length
Sequence Reads (Million)
18
16
14
12
10
8
6
4
2
58
59
60
Sequence Length (bp)

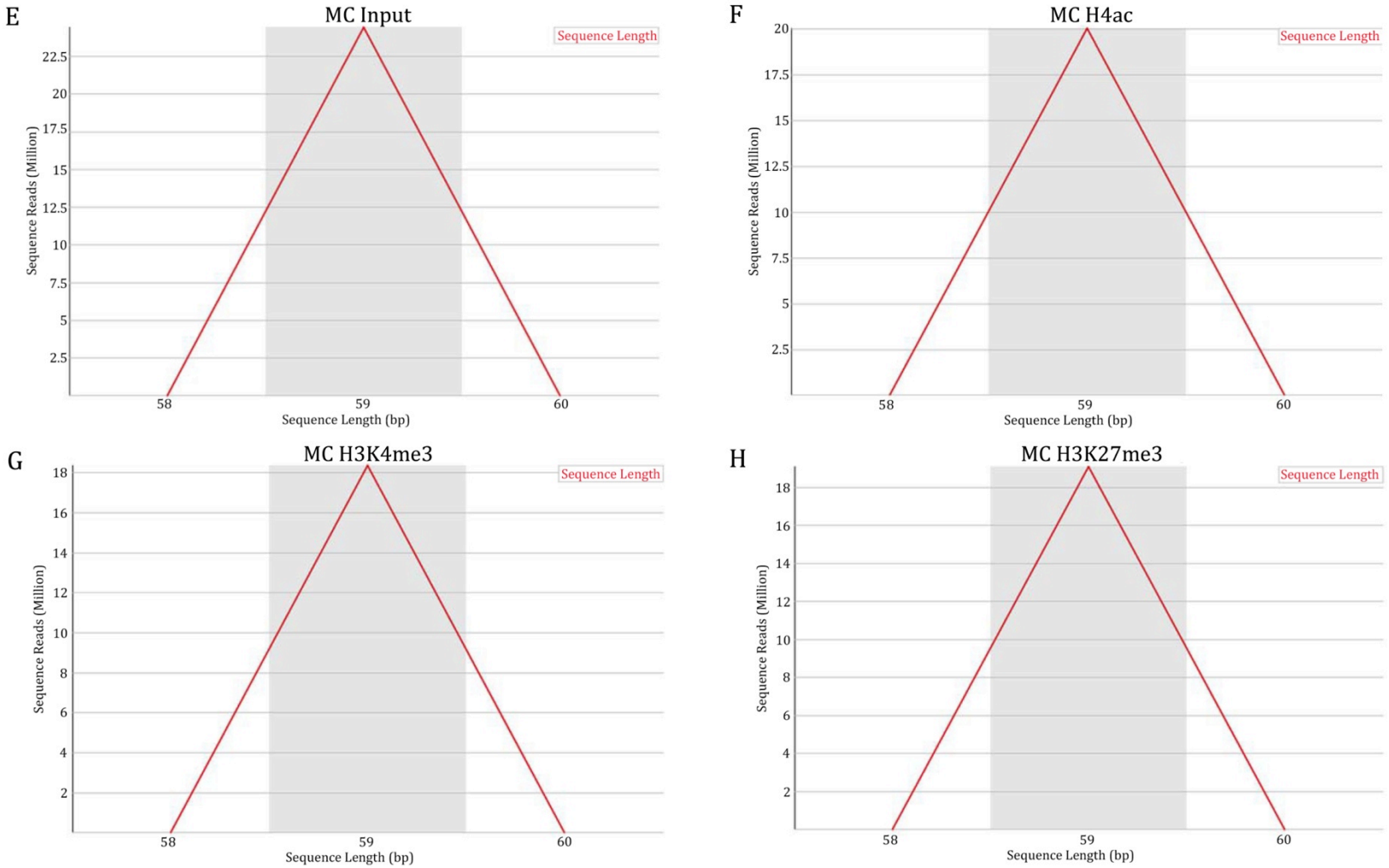



**Figure 2-8. Sequence length distribution report generated from FastQC analysis.** Distribution of sequence lengths over all sequences for Sertoli/germ (A-D) and myoid cell (E-F) Input (A, E), H4ac (B, F), H3K4me3 (C, G) and H3K27me3 (D, H).

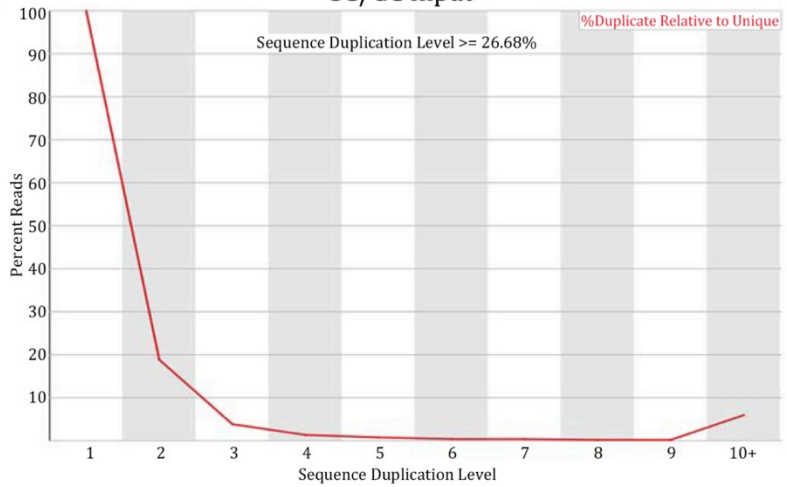
*Duplicate Sequences*

Comparing the first 50 bases across the first 200,000 sequence tags within each sample identifies duplicate sequences. Two sequence tags are considered identical if they are 100% identical across the first 50. A warning is issued for samples with duplicates greater than 20% of the total, and a failure issued for samples with duplicates greater than 50% of the total.

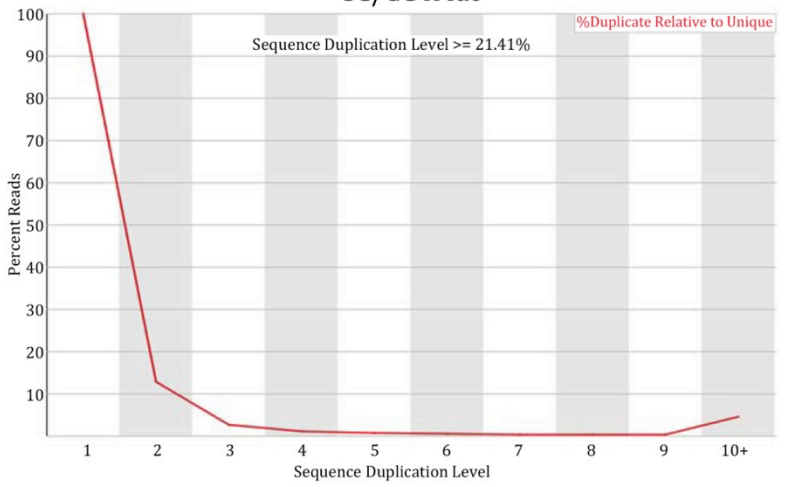
Duplicate sequence tags are thought to arise during the library preparation. During this step, ChIP DNA is intentionally amplified to increase the amount of DNA required for sequencing. Numerous duplicates suggest that insufficient starting material was used in preparing the library or that the DNA was not appropriately size fractionated prior to library preparation (smaller fragments will amplify more than larger fragments).

The number of duplicate sequence tags in each sample set was graphed as percentage of the total number of sequence tags (y-axis) versus the number of times that a specific sequence occurred in the library (x-axis). Only one ChIP sample, MC H4ac, passed this quality report while the others either failed or evoked a warning, indicating the need to remove duplicate sequence tags prior to alignment (Figure 2-9 A-H). The MC H4ac and MC H3K27me3 samples had the highest duplication rates, 51% and 71% respectively (Figure 2-9 F and H). This agrees with the per sequence GC% report discussed above that indicated there were biased sequences in the MC H3K27me3 sample. All duplicate sequence tags were removed from the samples prior to genome alignment.

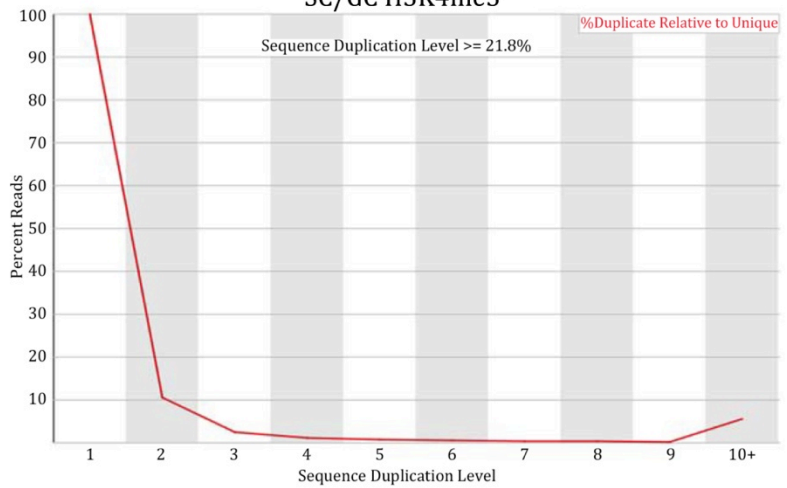
A


SC/GC Input
%Duplicate Relative to Unique
Sequence Duplication Level >= 26.68%
Percent Reads
100
90
80
70
60
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10+
Sequence Duplication Level

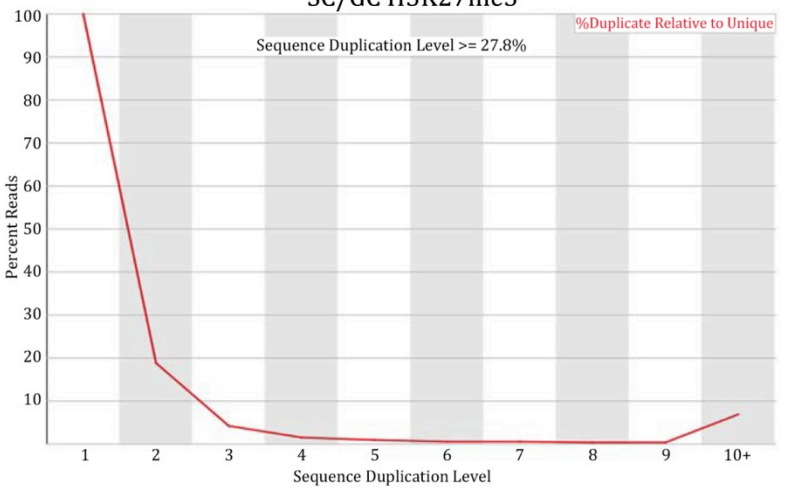
B


SC/GC H4ac
%Duplicate Relative to Unique
Sequence Duplication Level >= 21.41%
Percent Reads
100
90
80
70
60
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10+
Sequence Duplication Level

C


SC/GC H3K4me3
%Duplicate Relative to Unique
Sequence Duplication Level >= 21.8%
Percent Reads
100
90
80
70
60
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10+
Sequence Duplication Level

D


SC/GC H3K27me3
%Duplicate Relative to Unique
Sequence Duplication Level >= 27.8%
Percent Reads
100
90
80
70
60
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10+
Sequence Duplication Level

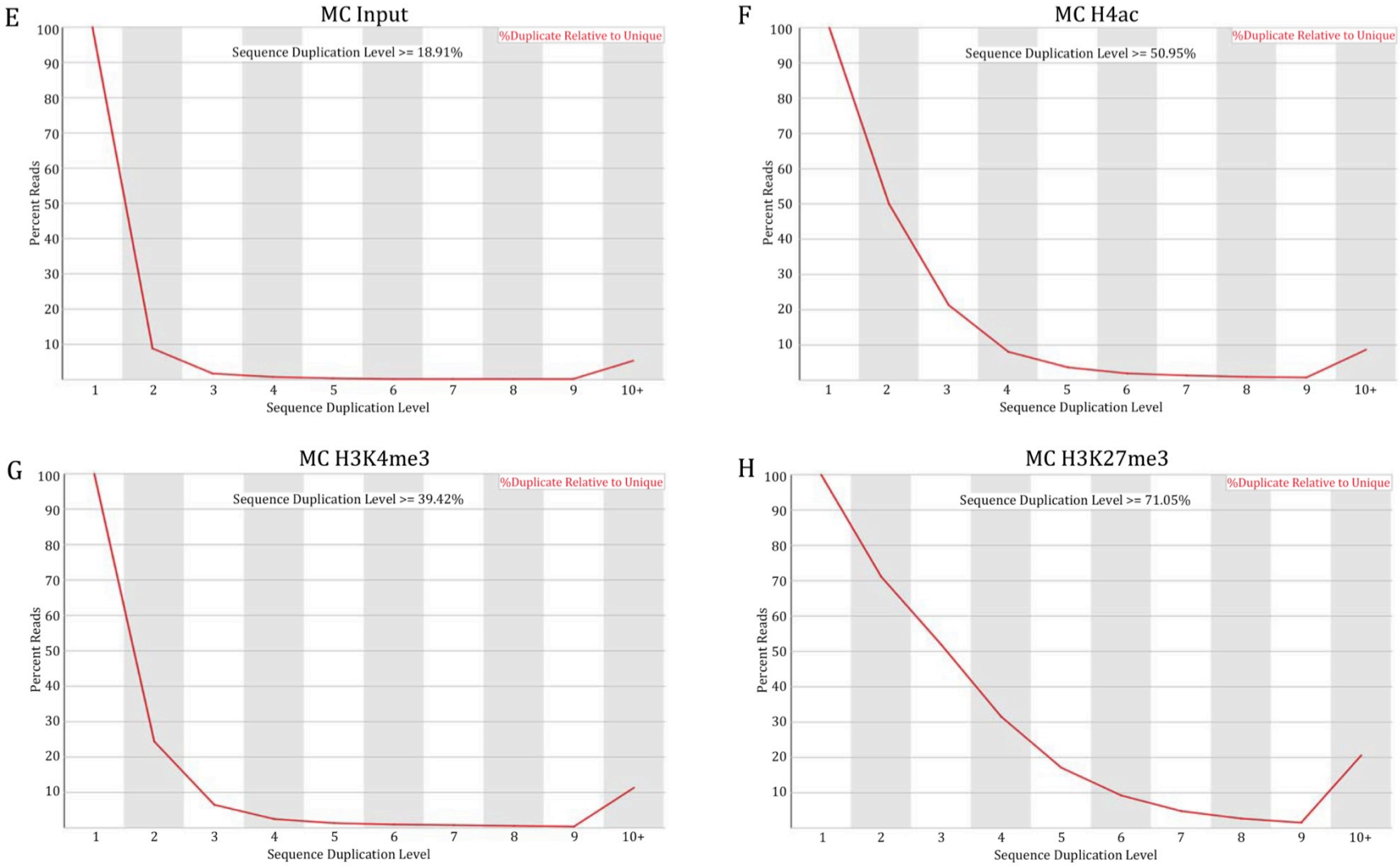



**Figure 2-9. Duplicate sequence report generated from FastQC analysis.** Sequence duplication levels for Sertoli/germ (A-D) and myoid cell (E-F) Input (A, E), H4ac (B, F), H3K4me3 (C, G) and H3K27me3 (D, H).

*Overrepresented Sequences*

This report identifies any linker DNA sequences that may present within sequence, i.e. adapter sequences and primer sequences. The first 200,000 sequence tags in each sample are analyzed and any similar sequences present in at least 0.1% of the total sequence tags are identified. Identified sequences are then cross-referenced with a database of known sequencing primers and adapters. Any adapter and primer sequences present need trimmed from each sequence tag prior to genome alignment.

Analysis of the data showed that only the input samples were free of linker DNA (Table 2-2). ChIP Sertoli/germ and myoid cell samples shared sequence identity similar to sequences in the database. A sequence similar to the Illumina Paired end primer 2 and the Illumina adapter sequence were identified, the primer sequence is underlined and the adapter sequence is italicized in Table 2-2. This primer/adapter sequence was ligated to all of the sequences prior to cluster amplification and sequencing. Two of the myoid cell samples, H4ac and H3K27me3, were also associated with a duplicate of the sequencing primer, this is likely an artifact that occurred during the ligation step and these sequence tags will be discarded.

**Table 2-2. Overrepresented sequence report generated from FastQC analysis.**

| | ChIP Sample | Overrepresented Sequence | Possible Source |
|---|---|---|---|
| Sertoli/ Germ | Input | None | |
| | H4ac | GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC<br>G*ATCTCGTATGCCGTCTTCTGCTT* | Illumina Paired End PCR Primer 2 (100% over 59bp) |
| | H3K4me3 | GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC<br>G*ATCTCGTATGCCGTCTTCTGCTT* | Illumina Paired End PCR Primer 2 (100% over 59bp) |
| | H3K27me3 | GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC<br>G*ATCTCGTATGCCGTCTTCTGCTT* | Illumina Paired End PCR Primer 2 (100% over 59bp) |
| Myoid | Input | None | |
| | H4ac | GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC<br>G*ATCTCGTATGCCGTCTTCTGCTT* | Illumina Paired End PCR Primer 2 (100% over 59bp) |
| | | GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGAT<br>CGGAAGAGCGGTTCAGCAGGAATGC | Illumina Paired End PCR Primer 2 (97% over 36bp) |
| | H3K4me3 | GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC<br>G*ATCTCGTATGCCGTCTTCTGCTT* | Illumina Paired End PCR Primer 2 (100% over 59bp) |
| | H3K27me3 | GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC<br>G*ATCTCGTATGCCGTCTTCTGCTT* | Illumina Paired End PCR Primer 2 (100% over 59bp) |
| | | GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGAT<br>CGGAAGAGCGGTTCAGCAGGAATGC | Illumina Paired End PCR Primer 2 (97% over 36bp) |

*Summary of FastQC analysis*

A summary of the FastQC reports is listed in Table 2-3. Overall, the reports identified two key issues with our sequence files that need to be addressed before the sequence files can be aligned to the reference genome. Each sequence tag needs the adapter and primer sequence trimmed, if present, and each sequence file needs to have all duplicate sequences removed.

**Table 2-3. Summary of FastQC analysis.**

| | Sertoli/Germ Cells | | | | Myoid Cells | | | |
|---|---|---|---|---|---|---|---|---|
| | H4Ac | H3K4me3 | H3K27me3 | Input | H4Ac | H3K4me3 | H3K27me3 | Input |
| Per Base Sequence Quality | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass |
| Per Sequence Quality | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass |
| Per Base Sequence Content | Warning | Fail | Warning | Fail | Fail | Fail | Fail | Fail |
| Per Base GC Content | Warning | Fail | Pass | Fail | Fail | Fail | Fail | Fail |
| Per Sequence GC Content | Pass | Fail | Pass | Pass | Fail | Warning | Warning | Pass |
| Per Base N Content | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass |
| Sequence Length Distribution | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass |
| Duplicate Sequences | Warning 21.41% | Warning 21.8% | Warning 27.8% | Warning 26.68% | Fail 50.95% | Warning 39.42% | Fail 71.05% | Pass 18.91% |

**Preparing Sequence Files for Alignment.** To prepare the sequence files for alignment, ligated primer/adapter sequences and duplicate sequences were removed. Primer/adapter sequence trimming was performed using Trim Galore!, a software program that removed the ligated sequence from each sequence tag within the sample sets. Sequence tags that were trimmed to <20 base pairs were removed from the sample set (Table 2-4). Duplicate sequence tags (tags sharing 100% identity across 59 base pairs) were removed from all sample sets using the using the 'rmdup' command in Samtools. As noted in the Duplicate Sequences report the MC H4ac and MC H3K27me3 had the highest number of duplicates to be removed (Figure 2-9 F and H).

**Table 2-4. ChIP-Seq trimming and duplicate removal.**

| | Sample | Total Raw Sequence Tags | Seq Tags Lost to Trimming | Duplicate Seq Tags Removed |
|---|---|---|---|---|
| **Sertoli/ Germ Cells** | Input | 21,384,126 | 316,649 | 1,483,654 |
| | H4ac | 15,285,827 | 340,378 | 174,026 |
| | H3K4me3 | 14,888,624 | 443,820 | 1,145,536 |
| | H3K27me3 | 19,192,554 | 951,948 | 407,132 |
| **Myoid Cells** | Input | 24,406,389 | 96,544 | 1,863,592 |
| | H4ac | 20,022,701 | 432,591 | 3,693,225 |
| | H3K4me3 | 18,352,272 | 245,562 | 2,915,671 |
| | H3K27me3 | 19,072,572 | 834,186 | 4,546,020 |

**Alignment to the rattus norvegicus genome.** After trimming and duplication removal, each ChIP-Seq sequence file was aligned to the rn5 build of the rat genome using the BWA alignment tool with default parameters. Two genome assemblies were used to create our libraries - Unmasked and Masked. The Unmasked genome contains all repetitive elements and is considered the 'complete' genome. The Masked genome does not contain known repetitive elements. These two distinctions are important when considering alignments because alignment programs have a difficult time aligning repetitive sequences. A sequence tag mapping to a repetitive region may align to several repetitive regions equally well and the program has no way of determining the correct alignment location. Thus, interpreting alignments and therefore enrichment patterns in repetitive regions is difficult. We have therefore aligned our samples using both genome assemblies to facilitate the identification of specific enriched regions (Chapter 3).

As expected, a higher percentage of sequence tags mapped to the unmasked genome (Table 2-5). For this alignment, 40-60% of the total sequence tags mapping to the reference genome is considered good (open communication forum:(Li, Schmieder et al. 2012)). All of our samples fall within this range except for two, the SC H4ac (11.5%) and SC H3K27me3 (21.5%) and will be discussed further below. A small percentage of sequence tags mapped to the Masked genome, as expected since this genome assembly is significantly smaller due to the exclusion of repetitive regions (Table 2-6). Again, only a very small portion of the SC H4a (3%) and SC H3K27me3 (7%) aligned to the reference genome.

The decreased percentage of aligned sequence tags in the SC H4ac and H3K27me3 samples could be remedied by trimming the sequence tags to include only the first 25 bases and allowing up to six mismatches. This would increase the percentage of mapped sequence tags to ~50%. Caution must be taken as shortening the sequence tag length and increasing the number of allowed mismatches severely decreases the alignment quality. For this reason, the alignments for these samples were kept using the original alignment parameters.

To determine if these samples were contaminated with other DNA, the samples were aligned against several genome assemblies (mouse, human, yeast, *E. coli*, *C. elegans* and common vectors). Neither sample aligned well to these genomes, the best alignment was to the mouse genome at ~4% aligning (Data not shown). To further determine why the sequence tags were not aligning, a blast search was preformed on ~1000 sequence tags that did not align to the reference genome. These results indicate that ~30% of the 1000 sequences tested aligned to various species of fish. It is possible that the salmon sperm blocked agarose beads used in the ChIP experiment carried over and contaminated the sequencing library.

**Table 2-5. ChIP-Seq reads aligned to the unmasked rn5 genome.**

| | Sample | Total Raw Reads | Total Reads for alignment | Mapped to + Strand | Mapped to - Strand | Total Mapped Reads | Total Unmapped Reads | % Mapped | % Unmapped |
|---|---|---|---|---|---|---|---|---|---|
| Sertoli/Germ Cells | Input | 21,384,126 | 19,583,823 | 5,736,180 | 5,740,784 | 11,476,964 | 8,106,859 | 58.60% | 41.40% |
| | H4ac | 15,285,827 | 14,771,423 | 847,329 | 844,831 | 1,692,160 | 13,079,263 | 11.46% | 88.54% |
| | H3K4me3 | 14,888,624 | 13,299,268 | 3,046,737 | 3,043,352 | 6,090,089 | 7,209,179 | 45.79% | 54.21% |
| | H3K27me3 | 19,192,554 | 17,833,474 | 1,916,010 | 1,917,517 | 3,833,527 | 13,999,947 | 21.50% | 78.50% |
| Myoid Cells | Input | 24,406,389 | 22,446,253 | 9,278,557 | 9,272,384 | 18,550,941 | 3,895,312 | 82.65% | 17.35% |
| | H4ac | 20,022,701 | 15,896,885 | 4,186,185 | 4,184,642 | 8,370,827 | 7,526,058 | 52.66% | 47.34% |
| | H3K4me3 | 18,352,272 | 15,191,039 | 5,572,907 | 5,571,093 | 11,144,000 | 4,047,039 | 73.36% | 26.64% |
| | H3K27me3 | 19,072,572 | 13,692,366 | 4,303,612 | 4,304,775 | 8,608,387 | 5,083,979 | 62.87% | 37.13% |

**Table 2-6. ChIP-Seq reads aligned to the masked rn5 genome.**

| | Sample | Total Raw Reads | Total Reads for alignment | Mapped to + Strand | Mapped to - Strand | Total Mapped Reads | Total Unmapped Reads | % Mapped | % Unmapped |
|---|---|---|---|---|---|---|---|---|---|
| Sertoli/Germ Cells | Input | 21,384,126 | 19,568,170 | 1,408,296 | 1,409,704 | 2,818,000 | 16,750,170 | 14.40% | 85.60% |
| | H4ac | 15,285,827 | 14,748,764 | 196,054 | 196,486 | 392,540 | 14,356,224 | 2.66% | 97.34% |
| | H3K4me3 | 14,888,624 | 13,617,314 | 1,119,220 | 1,121,313 | 2,240,533 | 11,376,781 | 16.45% | 83.55% |
| | H3K27me3 | 19,192,554 | 17,718,984 | 606,480 | 609,518 | 1,215,998 | 16,502,986 | 6.86% | 93.14% |
| Myoid Cells | Input | 24,406,389 | 22,232,001 | 2,674,580 | 2,673,969 | 5,348,549 | 16,883,452 | 24.06% | 75.94% |
| | H4ac | 20,022,701 | 16,672,547 | 2,536,329 | 2,535,398 | 5,071,727 | 11,600,820 | 30.42% | 69.58% |
| | H3K4me3 | 18,352,272 | 15,752,650 | 1,334,348 | 1,333,787 | 2,668,135 | 13,084,515 | 16.94% | 83.06% |
| | H3K27me3 | 19,072,572 | 14,733,029 | 1,084,883 | 1,087,804 | 2,172,687 | 12,560,342 | 14.75% | 85.25% |

**Annotation of histone modifications across the rat genome.** The aligned ChIP-Seq files were viewed by converting, the sam (.sam) file format to a BigWig (.bw) file format, which were imported to and viewed in the UCSC Genome Browser. The histone modifications visualized across Chromosome 2 of the *rattus norvegicus* genome, rn5 build and a smaller region of Chromosome 2 are shown in Figures 2-10 and 2-11, respectively. Chromosome 2 was to represent general patters associated with the entire genome. Several features are represented within the Genome Browser view. A scale bar is listed across the top of the chart for reference, known RefSeq Genes are identified with purple bars, and the histogram shows enrichment of sequence tags mapped to the masked genome. Enrichment in each location is reflected by the height of the histogram. As noted above, the Sertoli cell H4ac and H3K27me3 did not demonstrate significant enrichment as only 2.6% and 6.8% of sequence tags aligned to the reference genome. This low percentage of aligned sequence tags prevents detailed evaluation of any trends within these samples.

As expected, the input samples for each cell type analyzed show relatively low enrichment across the chromosome. The tall, slender peaks observed in the input samples, as well as other samples, likely correspond to regions of DNA that were inadvertently enriched by the imperfect nature of nuclease digestion. As these peaks were seen in all samples within the same genomic location across each cell type, they do not represent regions enriched for specific histone modifications. H4ac enrichment in the myoid cell sample followed a fluctuating pattern that appears to coordinate with gene locations (Figure 2-11). This histone modification is typically found in euchromatic regions and is correlated with active gene

expression, suggesting that genes in this region are active or can be readily activated. The H3K27me3 density in myoid cells followed a more consistent pattern across the chromosome, and this pattern was also similar across other chromosomes (data not shown). As enrichment peaks associated with H3K27me3 are relatively small, visual association and comparative analysis of enrichment with other histone modifications will require bioinformatic analysis. The H3K4me3 enrichment profile follows a similar, more site-specific profile across both cell types, and this pattern is also found across other chromosomes (data not shown). H3K4me3 enrichment was similar between Sertoli and myoid cells; however, several distinct cell-specific peaks could be identified (Figure 2-11). Within the myoid cell samples, H3K4me3 seems to follow enrichment of H4ac although there are regions of H4ac that do not contain H3K4me3 enrichment. This corresponds with the theory that H4ac is a broader mark identifying larger genomic regions whereas H3K4me3 tends to associate with transcriptional start sites and enhancer regions.

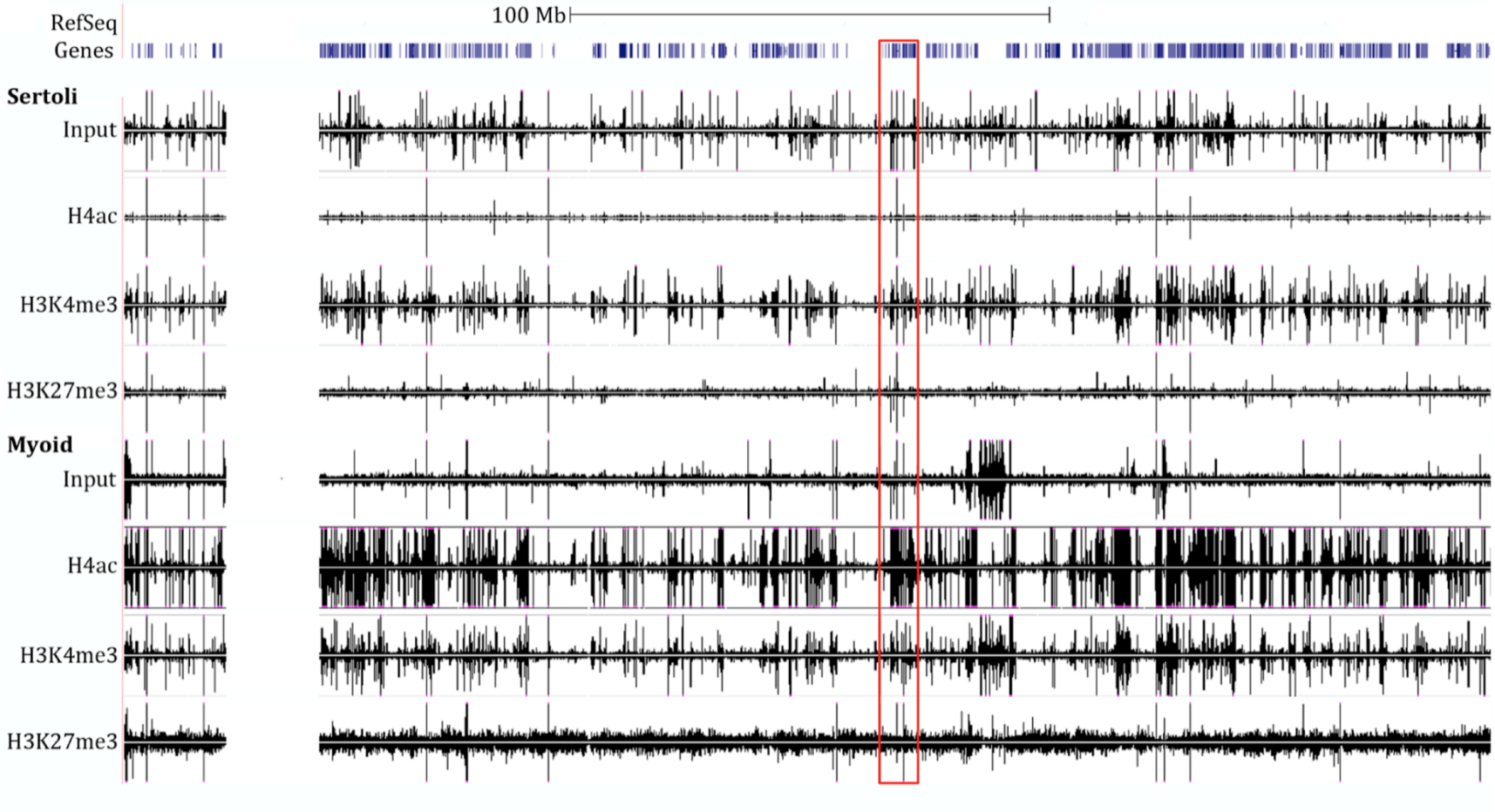



**Figure 2-10. Graphic representation of ChIP-Seq enrichment for each modification tested across Chromosome 2.** Sequence files aligned to the masked genome represented in the UCSC Genome Browser. Maximum enrichment values are plotted for each tag. Purple vertical bars represent gene locations. Red box indicates region in Figure 2-11.

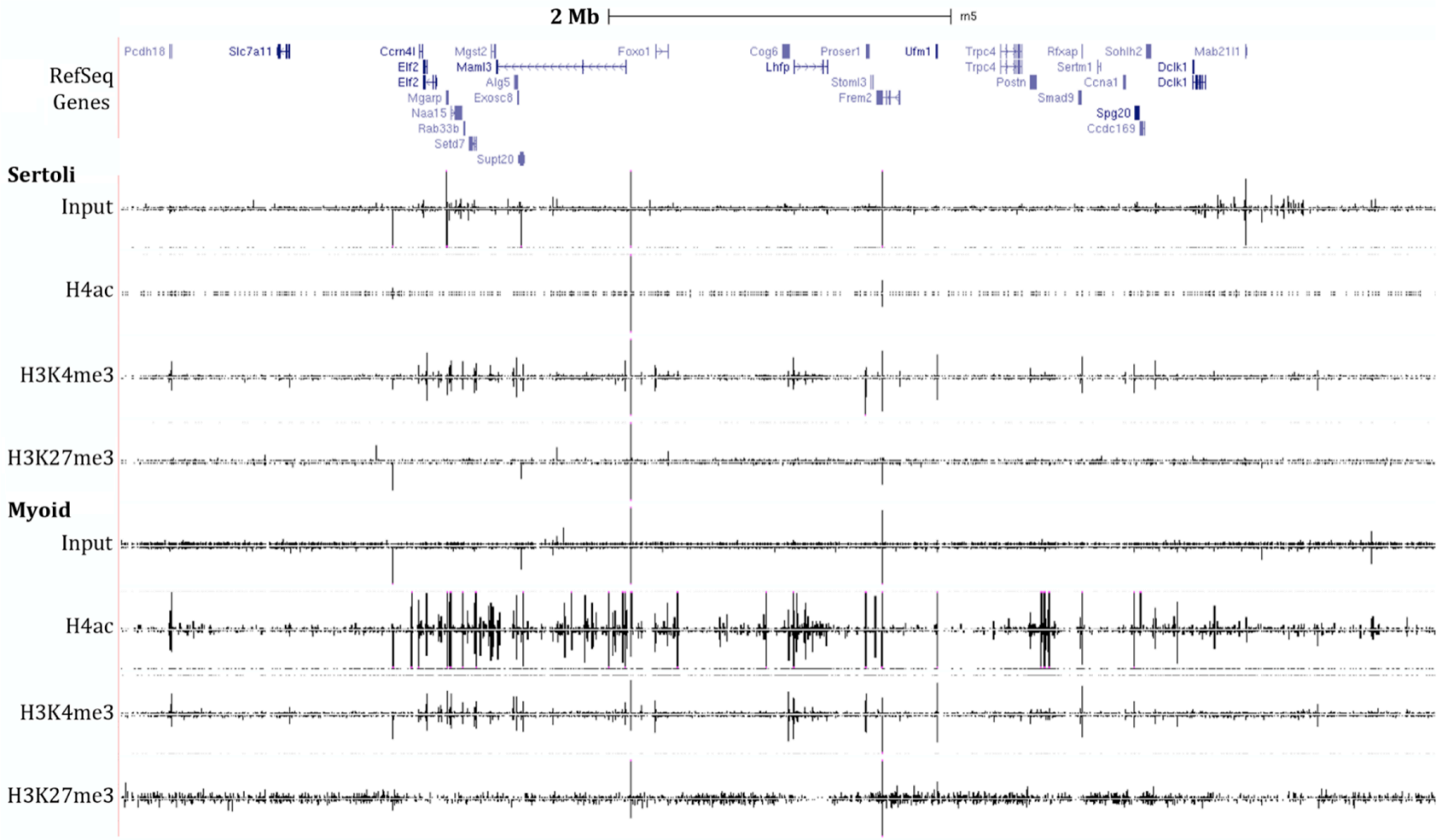



**Figure 2-11. Graphic representation of ChIP-Seq enrichment for each modification tested across a subset of Chromosome 2.** Sequence files aligned to the masked genome represented in the UCSC Genome Browser. Maximum enrichment values are plotted for each read tag. Purple vertical bars represent gene locations.

## Summary

Chromatin immunoprecipitation has allowed investigators to readily identify protein/DNA interactions *in vivo*, providing a snapshot view of DNA-bound protein interactions at specific loci at a specific time. When combined with next-generation sequencing capabilities, ChIP provides a genome-wide view of protein/DNA interactions.

Many factors contribute to the successful outcome of a ChIP-Seq experiment. As library preparation and sequencing are often outsourced, it is in the hands of the investigator to prepare the best possible DNA for sequencing. The initial step involves cross-linking the proteins to the DNA, followed by shearing the DNA. If the cells are over or under cross-linked, interactions may be falsely identified or relevant interactions may be missed completely. To ensure proper fixation in our cells, known protein/DNA interactions were assessed prior to sequencing. Fragmenting the DNA to the appropriate size (~200-600 bp for standard ChIP-PCR ~100-300 bp for ChIP-Seq) is a critical step that ensures proper identification of enrichment and, perhaps more importantly, the resolution of enrichment. For our studies we used micrococcal nuclease uniformly fragment the DNA to ~150 base pairs, and increase the resolution of our ChIP experiments. Another important factor contributing to the success of ChIP is antibody selection. For our studies, only antibodies known to work in ChIP assays were selected and enrichment for each antibody was tested using known regions of protein/DNA interaction prior to sequencing. The resulting ChIP-DNA must be eluted, washed and free of contaminating salts for accurate library preparation and sequencing.

Aside from requiring high quality ChIP-DNA and good quality sequencing, the ability to accurately align the generated sequence tags is the most difficult step facing whole genome analysis of protein/DNA interactions. It is often necessary to outsource the analysis of sequencing datasets to departments that specialize in the manipulation and analysis of such large data files. Numerous programs are now available for aligning raw sequence data, but often times a solid understanding of programming and computer language are required to accurately process and align such data sets.

Before alignment can proceed, raw sequence tags (often obtained as .fastq files) require processing, as failure to do so will result in poor genome alignments. Running the FastQC analysis software on raw sequence files identifies problems associated with each sequence file. Short sequence tags will often contain sequence belonging to the adapter and primer sequences used for library preparation, and need trimmed from all sequence tags within the data set. Several programs exist to accomplish this and we used Trim Galore! to trim our sequence tags.

Another processing step that must be performed before aligning the sequence tags to a reference genome is removal of duplicate sequence tags. Duplicate sequence tags are tags that contain identical start and stop coordinates, and arise within data sets as a PCR artifact during library preparation. Theoretically immunoprecipitated DNA fragments generated by shearing or micrococcal nuclease digestion will rarely lead to identical fragment isolation, thus duplicate sequence tags are removed to prevent the false identification of enriched regions. For this

study, the 'rmdup' tool included in the BWA alignment software package was used to remove identical sequences.

After these two steps are complete, the resulting sequence tags are ready for alignment. For our data sets, the Burrows-Wheeler Aligner (a.k.a. BWA) was used. This alignment program was designed to align short sequence tags with a balance between accuracy and performance (speed of alignment). This program automatically adjusts the alignment parameters' based on sequence tag length and detected error rate, aligning sequence tags with a <3% error only. The BWA aligner will also index the raw FASTQ sequence file prior to alignment without the need for additional software. Indexing basically converts the .fastq format into a tab delineated text (.txt) format that the alignment software can process.

All software programs available for aligning sequencing data will produce alignment files; however, the quality of the alignments must be considered. Short communications in sequencing blog sites imply that anywhere from 40-90% of the trimmed and de-duped sequence tags should align to the unmasked reference genome. The mouse and human genome are more complete and thus should generate better alignment scores. The rat genome, with the latest build released March 2012, is only about 90% complete. When aligning sequence tags to incomplete genomes, one cannot expect all sequence tags to align. Overall, most of our samples mapped well (~50-80% sequence tags mapping to the reference genome).

When aligning sequencing tags to masked genomes or specifically to promoter regions (genome files that only contain 1, 2 or 5 kb regions 5' of the

transcription start site) the accuracy of alignment increases at the as these regions usually fully sequenced. Thus the statistical power of the alignment increases although the overall percentage of mapped sequencing tags will decrease.

Two sequencing samples within our data sets did not align well to the reference genome. There are numerous reasons as to how this can occur including poor ChIP-DNA sample quality, poor library preparation and/or poor sequencing run. In our case, all ChIP-DNA was tested before sequencing by qPCR and each sample displayed an enrichment profile within *Fshr* promoter region that matched previous ChIP results. We used this as an indication that the ChIP assay worked correctly and yielded quality DNA. The library preparation and sequencing were all preformed by Cofactor Genomics and passed all of their quality analysis tests before and during the library preparation and sequencing process.

Several steps were taken to identify the why the SC-H4ac and SC-H3K27me3 samples had poor alignment scores. To determine if the unmapped sequencing tags were not aligned due to an increase in mismatched bases, the alignment parameters were relaxed to allow for more miss paired base calls. This did increase the number of aligned sequence tags, but the alignment quality could not be assured due to allowance of numerous mismatches (data not shown). To determine if the original ChIP-DNA was contaminated with another species of DNA, the samples were aligned to the human, mouse and yeast genomes. These alignments generated <5% mapped sequence tags for each genome (data not shown) indicating that other DNA used in our lab did not contaminate the ChIP samples. To determine the origin of the DNA within the unmapped sequence tags, random sets of sequences were entered into a

blastn search. Interestingly, ~30% of the sequences checked (~1000 sequence tags) mapped to various members of the Osteichthyes class. One possible source of this contamination is that the agarose beads used for the ChIP assay were blocked in salmon sperm DNA. However, this is the case for all of the samples. In short, these two samples will need re-sequenced before further anlaysis.

Overall, we have generated a reference database that maps H4ac, H3K4me3 and H3K27me3 enrichment in myoid cells and H3K4me3 enrichment in Sertoli cells across the *rattus norvigicus* genome. These genomic databases can be further mined to identify genomic regions and patterns associated with each enrichment profile.

# Chapter 3

## Genome-wide Analysis of Histone 3 Lysine 4 tri-methylation in P15 Sprague Dawley Sertoli Cells

## Abstract

Genome wide mapping of post-translational histone modifications identified specific patterns associated with chromatin states and gene activity which have facilitated identification of regulatory elements. As a means to better understand transcriptional regulation in Sertoli cells, we used ChIP-Seq to reveal regions across the genome enriched for Histone 3 Lysine 4 tri-methylation (H3K4me3), a modification associated with active promoter and enhancer elements. Bioinformatic analysis comparing regions enriched for H3K4me3 in both Sertoli and myoid cells was used to identify Sertoli cell specific enriched regions. Three genome assemblies were used for database generation: Unmasked, masked and a build containing only 1000 bp upstream of RefSeq genes. These three assemblies allow for the identification of enriched regions across the entire genome, in a genome that doesn't contain repetitive regions, and within upstream promoter regions respectively. Within Sertoli cells 37,371, 26,804 and 2,372 enriched regions were identified using the unmasked, masked and promoter only genome assemblies respectively. Within myoid cells 59,441, 27,509 and 2,015 enriched regions were identified using the unmasked, masked and promoter only genome assemblies respectively. Enrichment in the unmasked myoid cell database was roughly 1.6 times higher than that in Sertoli cells and may be attributed to the fact that myoid cells are largely undifferentiated compared to Sertoli cells. Of the identified regions 22,914, 9,777 and 944 respectively were identified as Sertoli cell specific for each genome assembly. For both Sertoli and myoid cells just over half of the identified regions enriched for H3K4me3 lie outside the promoter region/Exon 1, which

contradicts several studies that identify this mark primarily with promoter regions. Only when looking at the Sertoli cell specific enriched regions was H3K4me3 enrichment primarily identified within promoter and intragenic regions. GEO expression analysis indicates that the top 40 Sertoli cell specific genes identified using the upstream promoter region were are correlated gene expression in E13 Sertoli cells. Overall, we have identified H3K4me3 enriched regions in Sertoli and myoid cells and generated a searchable database for future analysis. Our data follows similar enrichment patterns identified by other groups for other cell types tested indicating that H3K4me3 can be used as a mark to access active genomic regions.

## Introduction

Linking specific covalent histone modifications to a specific function has been the focus of numerous research projects over the last several years, and numerous sites have been identified within histone tails that can be post-translationally modified. This has led to the idea that a 'code' may exist where by specific combinations of modifications may recruit specific proteins and initiate a specific action; However, there are arguments for and against whether or not this type of 'code' exists *in vivo* (reviewed in (Rando 2012)). Overall, one theme does prevail – some modifications are consistently associated with a specific function across numerous species and cell types (O'Neill and Turner 1995; Roh, Ngau et al. 2004; Calestagne-Morelli and Ausio 2006; Akkers, van Heeringen et al. 2009; Pauler, Sloane et al. 2009; Young, Willson et al. 2011). As described in Chapter 2, we have generated a genome wide map of H3K4me3 in Sertoli and myoid cells, which was the focus of this investigation.

Enrichment of H3K4me3 is commonly associated with 'active' chromatin regions. Enrichment is found at the transcriptional start site (TSS) and within the promoter regions of active genes. H3K4me3 enrichment in combination with H3K27me3 enrichment is also found within the promoter regions of genes that are poised for transcriptional activation. More recently, H3K4me3 enrichment was associated with cell-type specific enhancers (Heintzman, Hon et al. 2009; Pekowska, Benoukraf et al. 2011; Tian, Jia et al. 2011). Therefore, H3K4me3 is a useful mark highlighting active regions within chromatin, as putative to promoters, regulatory elements and cell-specific enhancers.

To help identify active elements and promoters specific to Sertoli cells, we compared the genome wide H3K4me3 enrichment profiles of Sertoli cells to that of peritubular myoid cells. ChIP-Seq data was analyzed using SICER- a software program that identifies H3K4me3 enriched regions- to identify and compare enriched regions in Sertoli and myoid cells (Zang, Schones et al. 2009). Our analysis identified enriched regions based on the number of sequence reads across a defined region compared to a control (input) sample for both Sertoli and myoid cells. Both databases were then compared and a list was generated based on enrichment only found within the Sertoli cell population. These databases provide not only cell-type specific enrichment profiles, but also offer a comparison between differentiated (Sertoli) and undifferentiated (myoid) cell types.

To date, the profile of H3K4me3 within Sertoli and myoid cells remains unexplored. Cui *et. al.* have explored the H3K4me3 profile in whole testis preparations and discovered that ~74% of promoters were marked by H3K4me3 within the testis (Cui, Liu et al. 2012). They also identified 1611'testis specific' promoters marked by H3K4me3 that were not identified within the cerebrum (Cui, Liu et al. 2012). As their study did not differentiate between testicular cell types, it is difficult to compare our results to theirs. We did identify fewer enriched regions, which may be attributed to the smaller sample population. Of the top 40 H3K4me3 enriched regions analyzed for each of our genome assemblies, over half of the Sertoli and myoid cell regions corresponded to 5' flanking and Exon 1 gene regions. The other regions were associated with intergenic sequence. Only when looking at

Sertoli cell specific enriched regions did the number of 5' flanking/Exon1 regions associated with H3K4me3 increase.

## Materials and Methods

**Animal Use, Cell Isolation, Chromatin Immunoprecipitation and Genome Sequencing.** Animal use, cell isolation, ChIP-Seq and genome alignment were described in Chapter 2 of this text. It is important to note that *Rattus norvegicus* genome is only ~90% sequenced at this time (current build rn5 March 2012) and does not contain any sequence information for the Y chromosome.

**Peak finding analysis.** Differential identification of enriched domains was used to compare H3K4me3 enrichment in ChIP-Seq data obtained from Sertoli and myoid cells. SICER v1.1, a software package that employs the SICER clustering methodology, was used to identify enriched H3K4me3 regions within both Sertoli and myoid cells for three genome assemblies: Unmasked, Masked and Promoter (1000 bp upstream of TSS) (Zang, Schones et al. 2009). A redundancy threshold of 1 was used to eliminate all duplicate reads from the data set prior to peak calling and comparison. A window size of 200, fragment size of 150, and gap size of 200 was used, as recommended for H3K4me3 histone data sets by the SICER documentation. For statistical cutoffs, an E-value of 1000 and a false discovery rate (FDR) of 1E-2 were used. Each database was normalized to an input control sample prior to identifying enriched regions within the genome. For the input sample, enriched regions were identified using a random background model with a lenient *E*-value threshold, and then the ChIP read count and control read count for each enriched region was used to determine the significance of enrichment, with the control read count rescaled to account for the size difference in the control library and the ChIP

library (Zang, Schones et al. 2009). Subtractive analysis was used to determine enriched regions that were present in Sertoli cells but not in myoid cells masked and unmasked genome assemblies. Additionally, H3K4me3 ChIP-Seq reads were aligned to only promoter regions (1000 bp upstream of TSS) and a differential comparison of peak locations were performed. For this analysis the same parameters were used for both alignment, peak calling, and peak comparison as were used for the whole-genome analysis.

## Results

**Identification of H3K4me3 Enrichment.** To identify regions of H3K4me3 enrichment, whole genome alignments for Sertoli and myoid cells were imported into SICER v1.1 along with their corresponding Input controls. The program used the standard, published settings for H3K4me3 localization and the corresponding input samples to control for background. Alignments were performed on three assemblies of the *Rattus norvegicus* genome Build 5 including the unmasked genome, masked genome (which does not include repeat regions) and a promoter-based genome that contains 1000 bp upstream from transcriptional start sites (TSS) of RefSeq genes. SICER v1.1 was also used to identify the Sertoli cell specific peaks by comparing the Sertoli and myoid cell enrichment profiles.

The total number of enriched regions identified by SICER cluster analysis is shown in Table 3-1 for Sertoli and myoid cells. Total enrichment was similar between Sertoli and myoid cells for the masked genome (26,804 vs. 27,509 respectively) and 9,777 regions were identified as Sertoli cell specific. Enrichment was higher in the unmasked genome (37,371 vs. 59,441) with an increased number of regions identified in the Myoid cell population. 22,914 of these regions were identified as Sertoli cell specific. Similar enrichment was found between Sertoli and myoid cell populations within the upstream promoter region (2,372 vs. 2,015 respectively) and 944 of those peaks were specific to Sertoli cells. Enriched regions not classified as Sertoli cell specific are not necessarily regions that are myoid cell specific as the identified regions may be present in both cell types and thus not considered Sertoli cell specific.

More enriched regions were identified in the Unmasked genome assembly. The increased number of enriched regions found in this assembly is due to the inclusion of repetitive sequences. Repetitive elements are biologically relevant to the structure and function of the genome, so is important to include repetitive regions in genome wide analysis (Conley and Jordan 2010; Ward, Wilson et al. 2013). Our results indicate that the myoid cell population contains more enrichment in repetitive regions than the Sertoli cell population. Myoid cells are continuously dividing whereas Sertoli cells are quiescent, implying that the Sertoli cell population contains more heterochromatin and thus less activity within the repetitive regions of the genome. This could explain this disparity of enrichment seen between the Sertoli and myoid cells in the unmasked genome assembly.

**Table 3-1. Total H3K4me3 enriched regions identified using SICER.**

| Base Genome | Total Enriched Regions Identified | | |
|---|---|---|---|
| | Sertoli Cell | Myoid Cell | Sertoli Cell Enriched |
| Rn5 Unmasked | 37,371 | 59,441 | 22,914 |
| Rn5 Masked | 26,804 | 27,509 | 9,777 |
| Rn5 1000bp Upstream of TSS | 2,372 | 2,015 | 944 |

**Sertoli cell H3K4me3 enriched sites.** The top Sertoli cell enriched H3K4me3 regions were identified by comparing the number of reads associated with the H3K4me3 sample to that in the input control. A list of genomic locations for each genome build (unmasked, masked and 1000 bp upstream region) was generated, and sorted for significance based on the fold enrichment of sample reads vs. input reads. The top 40 enriched locations (based on highest fold change) are listed in Tables 3-2, 3-3 and 3-4 for the unmasked, masked and 1000 bp upstream region genome assemblies.

Table 3-2 identifies the location of peaks identified using the unmasked genome build for alignment (i.e. contains all repeat elements). Of the 40 regions, 24 were located within the 5' flanking region and first exon/intron of a gene, two were located within the gene body and 14 were located in intergenic regions. For the peaks located within intergenic regions, the closest 5' and 3' genes are listed (respectively). Table 3-3 identifies the top 40 enriched regions identified using the masked genome build for alignment. Of these, 24 were located within the 5' flanking region and first exon/intron a gene, two were located within the gene body and 14 were located in intergenic regions. This data suggests that H3K4me3 enrichment primarily centers around transcriptional start sites, but may also identify important intergenic regulatory regions.

When the regions identified using the unmasked genome were compared to those from the masked genome, 10 regions were identified in both lists – seven corresponding to promoter regions and three to intergenic regions (indicated Table 3-2 and 3-3 by italic type font). Identification of enriched regions across both

genome assemblies indicates a higher level of significance since the tables are ranked by fold enrichment. Regions not cross-identified in both genome assemblies does not mean that the region is not found within both assemblies- only the top 40 enriched regions were cross-compared and displayed here. Numerous regions are found in both datasets, but the level of significance is not the same between the two (data not shown).

Table 3-4 identifies Sertoli cell H3K4me3 enriched regions located 1000 bp upstream of known TSS. As whole genome expression analysis for P15 Sertoli cells is not currently available, gene expression was determined by searching the GEO database for whole testis expression at any age. All of the genes identified were associated with testis gene expression except one and no expression data was available for that gene (RGD1563714). This suggests that H3K4me3 enrichment can also identify transcriptionally active genes in Sertoli cells as in other tested cell types.

**Table 3-2. Top 40 H3K4me3 enriched regions identified in Sertoli cells – Unmasked genome assembly.**

| Chromosome Location | Size (bp) | SC Reads | Input Reads | P-Value | Fold Change | FDR | Gene Association | Location |
|---|---|---|---|---|---|---|---|---|
| chr18:47596400-47598599 | 2199 | 129 | 0 | 0 | 972.4 | 0 | Ppic | Promoter, Exon/Intron 1 |
| chr5:64168200-64170599 | 2399 | 123 | 0 | 0 | 927.2 | 0 | *Rnf38, Melk* | Intergenic |
| chr4:246732800-246735799 | 2999 | 122 | 0 | 0 | 919.6 | 0 | Ergic2 | Promoter, Exon/Intron 1 |
| chr10:14138000-14140599 | 2599 | 115 | 0 | 4.47E-293 | 866.9 | 5.85E-290 | *Hn1l* | Promoter, Exon/Intron 1 |
| chr2:147356800-147358799 | 1999 | 114 | 0 | 3.91E-290 | 859.3 | 4.56E-287 | *Mfsd8* | Intron/Exon 2-4 |
| chr3:99425000-99426999 | 1999 | 112 | 0 | 2.91E-284 | 844.3 | 2.89E-281 | Cd44 | Promoter, Exon/Intron 1 |
| chr12:54405800-54408399 | 2599 | 105 | 0 | 7.84E-264 | 791.5 | 4.79E-261 | Gtpbp6 | Promoter, Exon/Intron 1 |
| chr1:65362800-65364599 | 1799 | 104 | 0 | 6.26E-261 | 784.0 | 3.62E-258 | *Zfp40* | Promoter, Exon/Intron 1 |
| chr1:248782800-248784599 | 1799 | 100 | 0 | 2.32E-249 | 753.8 | 9.79E-247 | *Ptar1* | Promoter, Exon/Intron 1 |
| chr7:11084600-11086199 | 1599 | 97 | 0 | 9.94E-241 | 731.2 | 3.30E-238 | LOC100125368 | Promoter, Exon/Intron 1 |
| chr3:144553400-144555799 | 2399 | 91 | 0 | 1.38E-223 | 686.0 | 3.12E-221 | *Dstn, Snx5* | Intergenic |
| chr1:52105200-52108599 | 3399 | 88 | 0 | 4.46E-215 | 663.3 | 8.08E-213 | Qk | Intron 2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr2:52159400-52162199 | 2799 | 86 | 0 | 1.98E-209 | 648.3 | 3.12E-207 | Cd180, Srek1 | Intergenic |
| chr10:75376400-75377599 | 1199 | 85 | 0 | 1.30E-206 | 640.7 | 1.93E-204 | *Rad51c* | Promoter, Exon/Intron 1 |
| chr13:1940400-1942799 | 2399 | 85 | 0 | 1.30E-206 | 640.7 | 1.93E-204 | Cntnap5a, Cntnap5c | Intergenic |
| chr20:13413200-13414799 | 1599 | 84 | 0 | 8.44E-204 | 633.2 | 1.17E-201 | *RGD1303003* | Promoter, Exon/Intron 1-2 |
| chr4:62429600-62431199 | 1599 | 83 | 0 | 5.41E-201 | 625.7 | 7.08E-199 | *Nup205* | Promoter, Exon/Intron 1 |
| chr10:62368400-62370799 | 2399 | 83 | 0 | 5.41E-201 | 625.7 | 7.08E-199 | Git1, Coro6 | Intergenic |
| chr1:105165000-105166999 | 1999 | 82 | 0 | 3.42E-198 | 618.1 | 4.27E-196 | Zdhhc13 | Promoter, Exon/Intron 1 |
| chr13:54919000-54921399 | 2399 | 82 | 0 | 3.42E-198 | 618.1 | 4.27E-196 | Pik3c2b | Promoter, Exon/Intron 1 |
| chr18:62314200-62316199 | 1999 | 82 | 0 | 3.42E-198 | 618.1 | 4.27E-196 | Tubb6 | Promoter, Exon/Intron 1-2 |
| chr1:66412600-66413799 | 1199 | 79 | 0 | 8.08E-190 | 595.5 | 8.43E-188 | Slc27a5, RGD1561909 | Intergenic |
| chr10:73532000-73533599 | 1599 | 79 | 0 | 8.08E-190 | 595.5 | 8.43E-188 | Ints2 | Promoter |
| chr8:127340200-127341599 | 1399 | 77 | 0 | 2.90E-184 | 580.4 | 2.70E-182 | Exog | Exon/Intron 1 |
| chr10:13435000-13436999 | 1999 | 77 | 0 | 2.90E-184 | 580.4 | 2.70E-182 | Ccnf | Promoter, Exon/Intron 1 |
| chr16:54354200-54356199 | 1999 | 77 | 0 | 2.90E-184 | 580.4 | 2.70E-182 | Slc7a2, Mtmr7 | Intergenic |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr8:22457000-22458599 | 1599 | 75 | 0 | 9.91E-179 | 565.4 | 8.10E-177 | Slc44a2, Ilf3 | Intergenic |
| chr9:100100600-100101999 | 1399 | 75 | 0 | 9.91E-179 | 565.4 | 8.10E-177 | Mterfd2 | Promoter, Exon/Intron 1 |
| chr8:84786000-84788399 | 2399 | 74 | 0 | 5.68E-176 | 557.8 | 4.32E-174 | Gclc, Elov15 | Intergenic |
| chrX:12436600-12439999 | 3399 | 74 | 0 | 5.68E-176 | 557.8 | 4.32E-174 | *Atp6ap2, Bcor* | Intergenic |
| chr14:82975000-82977399 | 2399 | 73 | 0 | 3.21E-173 | 550.3 | 2.33E-171 | Fgfr3 | Promoter, Exon/Intron 1 |
| chr3:66159200-66160999 | 1799 | 72 | 0 | 1.79E-170 | 542.7 | 1.23E-168 | Cdca7, Ola1 | Intergenic |
| chr8:106649000-106650799 | 1799 | 72 | 0 | 1.79E-170 | 542.7 | 1.23E-168 | Faim | Promoter, Exon/Intron 1 |
| chr2:108114400-108116199 | 1799 | 71 | 0 | 9.85E-168 | 535.2 | 6.31E-166 | RGD1565641 | Promoter, Exon/Intron 1 |
| chr1:82415400-82416399 | 999 | 70 | 0 | 5.35E-165 | 527.7 | 3.24E-163 | Zfp94 | Promoter, Exon/Intron 1 |
| chr20:20852400-20855199 | 2799 | 70 | 0 | 5.35E-165 | 527.7 | 3.24E-163 | Tfam, Bicc1 | Intergenic |
| chr3:157839800-157841799 | 1999 | 69 | 0 | 2.86E-162 | 520.1 | 1.62E-160 | Uqcc | Promoter, Exon/Intron 1 |
| chr10:52051800-52053599 | 1799 | 69 | 0 | 2.86E-162 | 520.1 | 1.62E-160 | Map2k4, Zfp18 | Intergenic |
| chr10:110678600-110679999 | 1399 | 69 | 0 | 2.86E-162 | 520.1 | 1.62E-160 | Metrn1, Vom2r65 | Intergenic |
| chr19:62972400-62974199 | 1799 | 69 | 0 | 2.86E-162 | 520.1 | 1.62E-160 | Dnaaf1 | Promoter, Exon/Intron 1 |

**Table 3-3. Top 40 H3K4me3 enriched regions identified in Sertoli cells – Masked genome assembly.**

| Chromosome Location | Size (bp) | SC Reads | Input Reads | P-Value | Fold Change | FDR | Gene Association | Location |
|---|---|---|---|---|---|---|---|---|
| chr18:63297000-63300399 | 3399 | 139 | 0 | 0 | 698.6 | 0 | Fam210a | Promoter, Exon/Intron 1 |
| chr5:64167600-64170799 | 3199 | 127 | 0 | 3.72E-306 | 638.3 | 9.06E-303 | *Rnf28, Melk* | Intergenic |
| chr2:147355600-147358999 | 3399 | 118 | 0 | 5.27E-281 | 593.1 | 8.10E-278 | *Mfsd8* | Exon/Intron 2-3 |
| chr10:14137800-14140599 | 2799 | 116 | 0 | 1.87E-275 | 583.0 | 2.48E-272 | *Hn1l* | Promoter, Exon/Intron 1 |
| chr1:48279600-48281999 | 2399 | 106 | 0 | 6.22E-248 | 532.8 | 4.66E-245 | Gtf2h5 | Promoter, Exon/Intron 1 |
| chr8:1496200-1498599 | 2399 | 104 | 0 | 1.78E-242 | 522.7 | 1.05E-239 | Msantd4 | Promoter, Exon/Intron 1 |
| chr10:108084600-108086399 | 1799 | 104 | 0 | 1.78E-242 | 522.7 | 1.05E-239 | Eif4a3, Nptx1 | Intergenic |
| chr11:50327200-50328999 | 1799 | 104 | 0 | 1.78E-242 | 522.7 | 1.05E-239 | Cep97 | Promoter, Exon/Intron 1 |
| chr15:61051000-61053399 | 2399 | 104 | 0 | 1.78E-242 | 522.7 | 1.05E-239 | Zc3h13 | Promoter, Exon/Intron 1 |
| chr1:248782600-248784799 | 2199 | 103 | 0 | 9.41E-240 | 517.7 | 4.74E-237 | *Ptar1* | Promoter, Exon/Intron 1 |
| chr1:62592200-62593999 | 1799 | 102 | 0 | 4.92E-237 | 512.7 | 2.25E-234 | Zfp52, Vom1r22 | Intergenic |
| chr7:12769400-12772599 | 3199 | 102 | 0 | 4.92E-237 | 512.7 | 2.25E-234 | Arid3a | Promoter |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr18:15622600-15624399 | 1799 | 102 | 0 | 4.92E-237 | 512.7 | 2.25E-234 | Zfp397 | Promoter, Exon/Intron 1 |
| chr8:21598000-21599599 | 1599 | 98 | 0 | 3.33E-226 | 492.6 | 1.08E-223 | Zfp266 | Promoter, Exon/Intron 1-2 |
| chr11:73005800-73008999 | 3199 | 98 | 0 | 3.33E-226 | 492.6 | 1.08E-223 | Kalrn | Intron 29 |
| chr13:88727600-88730199 | 2599 | 98 | 0 | 3.33E-226 | 492.6 | 1.08E-223 | Mpzl1, Rcsd1 | Intergenic |
| chr5:154155600-154157599 | 1999 | 96 | 0 | 8.17E-221 | 482.5 | 2.27E-218 | Taf12, Secp43 | Intergenic |
| chrX:12435200-12439999 | 4799 | 95 | 0 | 3.98E-218 | 477.5 | 1.03E-215 | *Atp6ap2, Bcor* | Intergenic |
| chr12:1328400-1331399 | 2999 | 94 | 0 | 1.92E-215 | 472.5 | 4.80E-213 | Stard13, Rfc3 | Intergenic |
| chr13:57184200-57186399 | 2199 | 92 | 0 | 4.33E-210 | 462.4 | 8.85E-208 | Timm17a | Promoter, Exon/Intron 1 |
| chr5:151637200-151639999 | 2799 | 91 | 0 | 2.03E-207 | 457.4 | 3.62E-205 | Tmem39b | Promoter, Exon/Intron 1 |
| chr18:3174400-3176399 | 1999 | 91 | 0 | 2.03E-207 | 457.4 | 3.62E-205 | Rbbp8 | Promoter, Exon/Intron 1 |
| chr2:219248800-219249999 | 1199 | 90 | 0 | 9.37E-205 | 452.3 | 1.56E-202 | Olr390, Fmo5 | Intergenic |
| chr1:181315400-181316999 | 1599 | 89 | 0 | 4.28E-202 | 447.3 | 6.48E-200 | Akip1 | Promoter, Exon/Intron 1 |
| chr5:73151000-73152799 | 1799 | 89 | 0 | 4.28E-202 | 447.3 | 6.48E-200 | Smc2 | Promoter, Exon/Intron 1 |
| chr10:11829200-11831399 | 2199 | 89 | 0 | 4.28E-202 | 447.3 | 6.48E-200 | Marf1, Olr1365 | Intergenic |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr10:75376000-75377799 | 1799 | 87 | 0 | 8.67E-197 | 437.3 | 1.20E-194 | *Rad51c* | Promoter, Exon/Intron 1 |
| chr17:45797400-45801799 | 4399 | 87 | 0 | 8.67E-197 | 437.3 | 1.20E-194 | Abt1 | Promoter, Exon/Intron 1 |
| chr3:144553000-144555999 | 2999 | 85 | 0 | 1.68E-191 | 427.2 | 1.91E-189 | *Dstn, Snx5* | Intergenic |
| chr5:80294800-80297399 | 2599 | 85 | 0 | 1.68E-191 | 427.2 | 1.91E-189 | Dnajc25 | Promoter, Exon/Intron 1 |
| chr17:26130800-26132799 | 1999 | 85 | 0 | 1.68E-191 | 427.2 | 1.91E-189 | Sirt5, Rnf182 | Intergenic |
| chr17:58618000-58620799 | 2799 | 85 | 0 | 1.68E-191 | 427.2 | 1.91E-189 | Olr1654, Hist1h2ak | Intergenic |
| chr5:144481800-144483999 | 2199 | 84 | 0 | 7.25E-189 | 422.2 | 7.51E-187 | Mycl1 | Promoter, Exon/Intron 1 |
| chr5:171581600-171583599 | 1999 | 83 | 0 | 3.10E-186 | 417.2 | 3.01E-184 | Park7, Tnfrsf9 | Intergenic |
| chr5:139677200-139678999 | 1799 | 82 | 0 | 1.31E-183 | 412.1 | 1.18E-181 | Eif2b3 | Promoter, Exon/Intron 1 |
| chr1:65362600-65364599 | 1999 | 81 | 0 | 5.45E-181 | 407.1 | 4.50E-179 | *Zfp40* | Promoter, Exon/Intron 1 |
| chr4:62429600-62431199 | 1599 | 81 | 0 | 5.45E-181 | 407.1 | 4.50E-179 | *Nup205* | Promoter, Exon/Intron 1 |
| chr10:3547400-3549599 | 2199 | 81 | 0 | 5.45E-181 | 407.1 | 4.50E-179 | Snn, Litaf | Intergenic |
| chr14:113611400-113612999 | 1599 | 81 | 0 | 5.45E-181 | 407.1 | 4.50E-179 | Mtif2 | Promoter, Exon/Intron 1 |
| chr20:13413200-13414799 | 1599 | 81 | 0 | 5.45E-181 | 407.1 | 4.50E-179 | *RGD1303003* | Promoter, Exon/Intron 1-2 |

**Table 3-4. Top 40 H3K4me3 enriched regions identified in Sertoli cells – 1000 bp upstream genome assembly.**

| Locus ID (1000 bp upstream) | Size (bp) | SC Reads | Input Reads | P-value | Fold Change | FDR | Gene Association | GEO Testis Expression |
|---|---|---|---|---|---|---|---|---|
| NM_053675 | 799 | 94 | 0 | 2.69E-195 | 289.5 | 3.09E-192 | Spata2 | Yes |
| NM_001024743 | 599 | 94 | 0 | 2.69E-195 | 289.5 | 3.09E-192 | Ugp2 | Yes |
| NM_012828 | 799 | 92 | 0 | 2.28E-190 | 283.4 | 1.66E-187 | Cacnb3 | Yes |
| NM_031520 | 599 | 92 | 0 | 2.28E-190 | 283.4 | 1.66E-187 | Myh10 | Yes |
| NM_001100840 | 999 | 91 | 0 | 6.54E-188 | 280.3 | 3.75E-185 | Txndc12 | Yes |
| NM_001109515 | 399 | 85 | 0 | 2.87E-173 | 261.8 | 1.44E-170 | Slc25a28 | Yes |
| NM_133620 | 799 | 78 | 0 | 2.04E-156 | 240.3 | 9.11E-154 | Zhx1 | Yes |
| NM_001107230 | 799 | 77 | 0 | 4.98E-154 | 237.2 | 1.81E-151 | Eif4enif1 | Yes |
| NM_001100556 | 599 | 75 | 0 | 2.84E-149 | 231.0 | 7.58E-147 | Gtf2e1 | Yes |
| NM_001170426 | 599 | 72 | 0 | 3.50E-142 | 221.8 | 7.19E-140 | Med25 | Yes |
| NM_133601 | 199 | 72 | 0 | 3.50E-142 | 221.8 | 7.19E-140 | Cblb | Yes |
| NM_001107494 | 399 | 71 | 0 | 7.86E-140 | 218.7 | 1.47E-137 | B9d2 | Yes |
| NM_001033064 | 599 | 71 | 0 | 7.86E-140 | 218.7 | 1.47E-137 | Kazald1 | Yes |
| NM_053871 | 799 | 70 | 0 | 1.74E-137 | 215.6 | 2.91E-135 | Srp54 | Yes |
| NM_053588 | 599 | 70 | 0 | 1.74E-137 | 215.6 | 2.91E-135 | Rnf138 | Yes |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NM_001105718 | 599 | 70 | 0 | 1.74E-137 | 215.6 | 2.91E-135 | Vps35 | Yes |
| NM_001013051 | 599 | 69 | 0 | 3.81E-135 | 212.5 | 5.46E-133 | Sugt1 | Yes |
| NM_012514 | 599 | 68 | 0 | 8.23E-133 | 209.5 | 1.06E-130 | Brca1 | Yes |
| NM_001106619 | 399 | 67 | 0 | 1.75E-130 | 206.4 | 1.95E-128 | Usp5 | Yes |
| NM_001134575 | 799 | 66 | 0 | 3.66E-128 | 203.3 | 3.81E-126 | RGD1308106 | Yes |
| NM_001108842 | 599 | 66 | 0 | 3.66E-128 | 203.3 | 3.81E-126 | Arl6 | Yes |
| NM_001108200 | 599 | 65 | 0 | 7.56E-126 | 200.2 | 7.48E-124 | Tbcc | Yes |
| NM_001012050 | 799 | 65 | 0 | 7.56E-126 | 200.2 | 7.48E-124 | Fbxo8 | Yes |
| NM_013070 | 599 | 64 | 0 | 1.54E-123 | 197.1 | 1.38E-121 | Utrn | Yes |
| NM_001139465 | 799 | 64 | 0 | 1.54E-123 | 197.1 | 1.38E-121 | RGD1310313 | Yes |
| NM_001106138 | 599 | 62 | 0 | 6.07E-119 | 191.0 | 5.02E-117 | Psmg2 | Yes |
| NM_031151 | 599 | 62 | 0 | 6.07E-119 | 191.0 | 5.02E-117 | Mdh2 | Yes |
| NM_133544 | 399 | 61 | 0 | 1.18E-116 | 187.9 | 8.91E-115 | Taf5 | Yes |
| NM_001012189 | 599 | 60 | 0 | 2.25E-114 | 184.8 | 1.54E-112 | Lpcat3 | Yes |
| NM_001107198 | 199 | 60 | 0 | 2.25E-114 | 184.8 | 1.54E-112 | Exo1 | Yes |
| NM_001105805 | 799 | 60 | 0 | 2.25E-114 | 184.8 | 1.54E-112 | Txndc17 | Yes |
| NM_001013983 | 599 | 60 | 0 | 2.25E-114 | 184.8 | 1.54E-112 | Mlec | Yes |
| NM_198766 | 599 | 59 | 0 | 4.23E-112 | 181.7 | 2.59E-110 | Ccdc127 | Yes |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NM_019376 | 399 | 59 | 0 | 4.23E-112 | 181.7 | 2.59E-110 | Ywhag | Yes |
| NM_001191846 | 799 | 57 | 0 | 1.42E-107 | 175.6 | 7.74E-106 | Foxo1 | Yes |
| NM_001109473 | 599 | 57 | 0 | 1.42E-107 | 175.6 | 7.74E-106 | Gtf3c4 | Yes |
| NM_001009710 | 799 | 57 | 0 | 1.42E-107 | 175.6 | 7.74E-106 | Lrrc41 | Yes |
| NM_001126297 | 599 | 57 | 0 | 1.42E-107 | 175.6 | 7.74E-106 | RGD1563714 | Unknown |
| NM_001024971 | 599 | 57 | 0 | 1.42E-107 | 175.6 | 7.74E-106 | Haus8 | Yes |
| NM_212549 | 599 | 57 | 0 | 1.42E-107 | 175.6 | 7.74E-106 | Ring1 | Yes |

**Myoid cell H3K4me3 enriched sites.** The top myoid cell enriched H3K4me3 regions were identified by comparing the number of reads associated with the H3K4me3 sample to that in the input control. A list of genomic locations for each genome build (unmasked, masked and 1000 bp upstream region) was generated, and sorted for significance based on the fold enrichment of sample reads vs. input reads. The top 40 enriched locations (based on highest fold change) are listed in Tables 3-5, 3-6 and 3-7 for the unmasked, masked and 1000 bp upstream region genome assemblies.

Table 3-5 identifies the location of peaks identified using the unmasked genome build for alignment. Of the 40 regions, 28 were located within the 5' flanking region and first exon/intron of a gene, three were located within the gene body and nine were located in intergenic regions. Table 3-6 identifies the top 40 enriched regions identified using the masked genome build for alignment. Of these, 30 were located within the 5' flanking region and first exon/intron a gene and 10 were located in intergenic regions. This data suggests that H3K4me3 enrichment primarily centers on transcriptional start sites, but may also identify important intergenic regulatory regions.

When the regions identified using the unmasked genome were compared to those from the masked genome, 13 regions were identified in both lists – nine corresponding to promoter regions and four to intergenic regions (indicated Table 3-5 and 3-6 by italic type font). Identification of enriched regions across both genome assemblies indicates a higher level of significance since the tables are ranked by fold enrichment. Regions not cross-identified in both genome assemblies

does not mean that the region is not found within both assemblies- only the top 40 enriched regions were cross-compared and displayed here. Numerous regions are found in both datasets, but the level of significance is not the same between the two (data not shown).

Table 3-7 identifies myoid cell H3K4me3 enriched regions located 1000 bp upstream of known TSS. As whole genome expression analysis for myoid cells is not currently available, gene expression was determined by searching the GEO database for whole testis expression at any age. All of the genes identified were associated with testis gene expression. This suggests that H3K4me3 enrichment can also identify transcriptionally active genes in myoid cells as in other tested cell types.

**Table 3-5. Top 40 H3K4me3 enriched regions identified in myoid cells – Unmasked genome assembly.**

| Chromosome Location | Size (bp) | SC Reads | Input Reads | P-Value | Fold Change | FDR | Gene Association | Location |
|---|---|---|---|---|---|---|---|---|
| chr7:120338400-120340199 | 1799 | 94 | 0 | 4.98E-227 | 625.9 | 5.27E-224 | *Eif3s6pi, Polr2f* | Intergenic |
| chr1:93896800-93898199 | 1399 | 79 | 0 | 1.62E-185 | 526.1 | 5.83E-183 | Tshz3 | Promoter |
| chr16:17376000-17377399 | 1399 | 74 | 0 | 6.10E-172 | 492.8 | 1.63E-169 | *RGD1309676* | Promoter, Exon/Intron 1 |
| chr5:144482400-144483599 | 1199 | 68 | 0 | 7.71E-156 | 452.8 | 1.36E-153 | *Mycl1* | Promoter |
| chr5:170370000-170371199 | 1199 | 64 | 0 | 3.15E-145 | 426.2 | 4.19E-143 | *Slc25a33* | Exon/Intron 1 |
| chr9:60526400-60527799 | 1399 | 64 | 0 | 3.15E-145 | 426.2 | 4.19E-143 | Hecw2, Ccdc150 | Intergenic |
| chr9:68228200-68229999 | 1799 | 62 | 0 | 5.81E-140 | 412.9 | 6.68E-138 | *Pard3b* | Promoter, Exon/Intron 1 |
| chr9:82511400-82512399 | 999 | 62 | 0 | 5.81E-140 | 412.9 | 6.68E-138 | *Slc4a3* | Promoter |
| chr1:267369000-267371199 | 2199 | 60 | 0 | 1.01E-134 | 399.5 | 9.90E-133 | Ccnj | Promoter, Exon/Intron 1-2 |
| chr10:107745600-107746599 | 999 | 60 | 0 | 1.01E-134 | 399.5 | 9.90E-133 | Cbx2 | Promoter, Exon/Intron 1-2 |
| chr1:190864600-190866799 | 2199 | 57 | 0 | 6.41E-127 | 379.6 | 4.95E-125 | Pde3b | Promoter, Exon/Intron 1 |
| chr3:19487000-19488799 | 1799 | 57 | 0 | 6.41E-127 | 379.6 | 4.95E-125 | LOC687813, Rab14 | Intergenic |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr5:161274000-161275199 | 1199 | 57 | 0 | 6.41E-127 | 379.6 | 4.95E-125 | Nbl1 | Exon/Intron 1 |
| chr4:7365400-7367399 | 1999 | 55 | 0 | 9.40E-122 | 366.2 | 6.29E-120 | Kcnh2 | Promoter |
| chr4:164988400-164989599 | 1199 | 55 | 0 | 9.40E-122 | 366.2 | 6.29E-120 | Tmem150a | Promoter |
| chr5:66150600-66152199 | 1599 | 55 | 0 | 9.40E-122 | 366.2 | 6.29E-120 | Tmod1 | Promoter, Exon/Intron 1 |
| chr15:63013000-63014799 | 1799 | 54 | 0 | 3.51E-119 | 359.6 | 2.17E-117 | *Serp2, Dnajc15* | Intergenic |
| chr17:11939400-11940599 | 1199 | 54 | 0 | 3.51E-119 | 359.6 | 2.17E-117 | *Mxd3* | Promoter, Exon/Intron 1-2 |
| chr1:173561400-173561999 | 599 | 53 | 0 | 1.28E-116 | 352.9 | 7.34E-115 | Stim1 | Upstream Promoter |
| chr10:86310400-86311199 | 799 | 53 | 0 | 1.28E-116 | 352.9 | 7.34E-115 | Zpbp2 | Promoter, Exon/Intron 1-2 |
| chr11:87046600-87048199 | 1599 | 53 | 0 | 1.28E-116 | 352.9 | 7.34E-115 | Camk2n2 | Promoter, Exon/Intron 1 |
| chr18:53441000-53442599 | 1599 | 53 | 0 | 1.28E-116 | 352.9 | 7.34E-115 | RGD1312005, MGC108823 | Intergenic |
| chr8:14330000-14331399 | 1399 | 52 | 0 | 4.62E-114 | 346.3 | 2.44E-112 | Slc36a4 | Promoter, Exon/Intron 1 |
| chr13:103926400-103927799 | 1399 | 52 | 0 | 4.62E-114 | 346.3 | 2.44E-112 | *Parp1, Mixl1* | Intergenic |
| chr15:42179200-42181199 | 1999 | 52 | 0 | 4.62E-114 | 346.3 | 2.44E-112 | *Fgf9* | Promoter, Exon/Intron 1 |
| chr8:62499000-62500799 | 1799 | 51 | 0 | 1.63E-111 | 339.6 | 7.97E-110 | *Sema7a* | Promoter, Exon/Intron 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr1:105630000-105631399 | 1399 | 50 | 0 | 5.64E-109 | 332.9 | 2.59E-107 | Nav2 | Upstream Promoter |
| chr8:28226600-28227799 | 1199 | 50 | 0 | 5.64E-109 | 332.9 | 2.59E-107 | Jam3 | Promoter, Exon/Intron 1 |
| chr8:116118600-116119599 | 999 | 50 | 0 | 5.64E-109 | 332.9 | 2.59E-107 | *Fam212a* | Promoter, Exon/Intron 1 |
| chr12:23780600-23782999 | 2399 | 48 | 0 | 6.38E-104 | 319.6 | 2.57E-102 | LOC100362783, Cldn3 | Intergenic |
| chr1:273330200-273331399 | 1199 | 47 | 0 | 2.08E-101 | 313.0 | 7.85E-100 | Pitx3 | Promoter |
| chr3:88003600-88004399 | 799 | 47 | 0 | 2.08E-101 | 313.0 | 7.85E-100 | Mapk8ip1 | Intron 2-3 |
| chr16:5730200-5731399 | 1199 | 47 | 0 | 2.08E-101 | 313.0 | 7.85E-100 | Cacna2d3 | Promoter, Exon/Intron 1 |
| chr9:3458400-3459599 | 1199 | 46 | 0 | 6.66E-99 | 306.3 | 2.35E-97 | Sult1c2a | Intron 3 |
| chr18:443800-444799 | 999 | 46 | 0 | 6.66E-99 | 306.3 | 2.35E-97 | F8 | Promoter, Exon/Intron 1 |
| chr19:66302600-66303399 | 799 | 46 | 0 | 6.66E-99 | 306.3 | 2.35E-97 | Cbfa2t3, Cdh15 | Intergenic |
| chr1:124237200-124238199 | 999 | 45 | 0 | 2.08E-96 | 299.7 | 6.90E-95 | Peg12 | Intron 1 |
| chr13:102290400-102291599 | 1199 | 45 | 0 | 2.08E-96 | 299.7 | 6.90E-95 | *Kif26b, Mixl1* | Intergenic |
| chr1:227976200-227977199 | 999 | 44 | 0 | 6.38E-94 | 293.0 | 1.97E-92 | Map3k11 | Upstream Promoter |
| chr1:195513400-195515399 | 1999 | 43 | 0 | 1.91E-91 | 286.3 | 5.57E-90 | RGD1565096 | Promoter, Exon/Intron 1 |

**Table 3-6. Top 40 H3K4me3 enriched regions identified in myoid cells – Masked genome assembly.**

| Chromosome Location | Size (bp) | SC Reads | Input Reads | P-Value | Fold Change | FDR | Gene Association | Location |
|---|---|---|---|---|---|---|---|---|
| chr7:120338400-120340799 | 2399 | 99 | 0 | 3.75E-249 | 793.7 | 4.47E-246 | *Eif3s6ip, RGD1359634* | Intergenic |
| chr9:109898000-109900999 | 2999 | 82 | 0 | 2.07E-200 | 657.4 | 5.37E-198 | Efna5 | Upstream Promoter |
| chr15:38289800-38290399 | 599 | 81 | 0 | 1.38E-197 | 649.4 | 3.16E-195 | Rabggta | Exon/Intron 1-2 |
| chr7:129499800-129501999 | 2199 | 78 | 0 | 3.78E-189 | 625.3 | 6.89E-187 | Alg12, Creld2 | Promoter, Exon/Intron 1-2 |
| chr5:144482400-144483799 | 1399 | 70 | 0 | 6.78E-167 | 561.2 | 7.11E-165 | *Mycl1* | Promoter, Intron 1 |
| chr8:118347800-118350199 | 2399 | 67 | 0 | 1.20E-158 | 537.2 | 9.92E-157 | Pth1r | Promoter, Exon/Intron 1 |
| chr1:227755600-227756999 | 1399 | 65 | 0 | 3.51E-153 | 521.1 | 2.42E-151 | Fosl1 | Promoter, Exon/Intron 1 |
| chr5:170370000-170371199 | 1199 | 64 | 0 | 1.86E-150 | 513.1 | 1.19E-148 | *Slc25a33* | Exon/Intron 1 |
| chr4:226890200-226891199 | 999 | 62 | 0 | 4.97E-145 | 497.1 | 2.69E-143 | MGC94282 | Promoter, Exon/Intron 1 |
| chr5:151471200-151472999 | 1799 | 62 | 0 | 4.97E-145 | 497.1 | 2.69E-143 | Hdac1 | Promoter, Exon/Intron 1 |
| chr2:219248800-219249999 | 1199 | 61 | 0 | 2.51E-142 | 489.1 | 1.26E-140 | Olr390, Fmo5 | Intergenic |
| chr2:215476400-215477399 | 999 | 60 | 0 | 1.25E-139 | 481.0 | 5.74E-138 | Tuft1, Pogz | Intergenic |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr16:17376200-17377399 | 1199 | 60 | 0 | 1.25E-139 | 481.0 | 5.74E-138 | *RGD1309676* | Promoter, Exon/Intron 1 |
| chr1:229458600-229459399 | 799 | 59 | 0 | 6.10E-137 | 473.0 | 2.56E-135 | Cox8a | Exon/Intron 1 |
| chr9:82511400-82512399 | 999 | 59 | 0 | 6.10E-137 | 473.0 | 2.56E-135 | *Slc4a3* | Promoter, Exon/Intron 1 |
| chr19:54884600-54885199 | 599 | 58 | 0 | 2.93E-134 | 465.0 | 1.14E-132 | Cfdp1 | Exon/Intron 1 |
| chr1:184027000-184028999 | 1999 | 57 | 0 | 1.39E-131 | 457.0 | 5.00E-130 | Dkk3, Mical2 | Intergenic |
| chr3:167089200-167089599 | 399 | 57 | 0 | 1.39E-131 | 457.0 | 5.00E-130 | Sdc4, Dbndd2 | Intergenic |
| chr9:68228200-68229999 | 1799 | 57 | 0 | 1.39E-131 | 457.0 | 5.00E-130 | *Pard3b* | Promoter, Exon/Intron 1 |
| chr13:103926400-103927799 | 1399 | 57 | 0 | 1.39E-131 | 457.0 | 5.00E-130 | *Parp1, Mixl1* | Intergenic |
| chr1:221344800-221346199 | 1399 | 54 | 0 | 1.32E-123 | 432.9 | 3.85E-122 | Rplp2 | Promoter, Exon/Intron 1-2 |
| chr4:7365400-7367399 | 1999 | 54 | 0 | 1.32E-123 | 432.9 | 3.85E-122 | Kcnh2 | Promoter, Exon/Intron 1 |
| chr15:42179200-42181399 | 2199 | 54 | 0 | 1.32E-123 | 432.9 | 3.85E-122 | *Fgf9* | Intron 1 |
| chr3:10813400-10814799 | 1399 | 53 | 0 | 5.84E-121 | 424.9 | 1.57E-119 | Surf6 | Promoter, Exon/Intron 1 |
| chr8:22874600-22875799 | 1199 | 53 | 0 | 5.84E-121 | 424.9 | 1.57E-119 | Kank2 | Promoter, Exon/Intron 1 |
| chr3:12984000-12984999 | 999 | 51 | 0 | 1.07E-115 | 408.9 | 2.48E-114 | RGD1564114, Ntng2 | Intergenic |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr15:63013000-63014799 | 1799 | 50 | 0 | 4.48E-113 | 400.9 | 9.62E-112 | *Serp2, Dnajc15* | Intergenic |
| chr17:11939400-11940599 | 1199 | 50 | 0 | 4.48E-113 | 400.9 | 9.62E-112 | *Mxd3* | Promoter, Exon/Intron 1-2 |
| chr12:9541000-9542199 | 1199 | 49 | 0 | 1.83E-110 | 392.8 | 3.64E-109 | Katnal1, Katnal1 | Intergenic |
| chr12:50380200-50381599 | 1399 | 49 | 0 | 1.83E-110 | 392.8 | 3.64E-109 | Tmem119 | Promoter, Exon/Intron 1 |
| chr1:220933200-220935199 | 1999 | 48 | 0 | 7.34E-108 | 384.8 | 1.37E-106 | Polr2l, Tspan4 | Promoter, Exon/Intron 1 |
| chr5:161274000-161275199 | 1199 | 48 | 0 | 7.34E-108 | 384.8 | 1.37E-106 | *Nbl1* | Promoter, Exon/Intron 1 |
| chr8:62498800-62500799 | 1999 | 48 | 0 | 7.34E-108 | 384.8 | 1.37E-106 | *Sema7a* | Promoter, Exon/Intron 1 |
| chr8:116118400-116119799 | 1399 | 48 | 0 | 7.34E-108 | 384.8 | 1.37E-106 | *Fam212a* | Promoter, Exon/Intron 1 |
| chr3:170399200-170401399 | 2199 | 47 | 0 | 2.88E-105 | 376.8 | 5.02E-104 | Rnf114 | Promoter, Exon/Intron 1 |
| chr10:107745600-107746599 | 999 | 47 | 0 | 2.88E-105 | 376.8 | 5.02E-104 | Cbx2 | Promoter, Exon/Intron 1-2 |
| chr13:102290200-102292399 | 2199 | 47 | 0 | 2.88E-105 | 376.8 | 5.02E-104 | *Kif26b, Mixl1* | Intergenic |
| chr5:172771000-172771799 | 799 | 46 | 0 | 1.11E-102 | 368.8 | 1.78E-101 | Tas1r1 | Upstream Promoter |
| chr17:64035200-64038999 | 3799 | 46 | 0 | 1.11E-102 | 368.8 | 1.78E-101 | Fzd8 | Exon 1 |
| chrX:45419000-45420799 | 1799 | 46 | 0 | 1.11E-102 | 368.8 | 1.78E-101 | Prkx | Promoter, Exon/Intron 1 |

**Table 3-7. Top 40 H3K4me3 enriched regions identified in myoid cells – Promoter genome assembly.**

| Chromosome Location | Size (bp) | SC Reads | Input Reads | P-Value | Fold Change | FDR | Gene Association | GEO Testis Expression |
|---|---|---|---|---|---|---|---|---|
| NM_001173433 | 799 | 98 | 0 | 1.99E-208 | 325.4 | 7.30E-205 | Tle1 | Yes |
| NM_001126268 | 799 | 94 | 0 | 2.18E-198 | 312.2 | 3.21E-195 | Tigd5 | Yes |
| NM_031339 | 799 | 94 | 0 | 2.18E-198 | 312.2 | 3.21E-195 | Parg | Yes |
| NM_001100518 | 599 | 77 | 0 | 1.45E-156 | 255.7 | 1.06E-153 | Hs2st1 | Yes |
| NM_001100698 | 799 | 77 | 0 | 1.45E-156 | 255.7 | 1.06E-153 | Zfp131 | Yes |
| NM_001024991 | 399 | 77 | 0 | 1.45E-156 | 255.7 | 1.06E-153 | Fahd1 | Yes |
| NM_031520 | 599 | 75 | 0 | 9.58E-152 | 249.1 | 5.03E-149 | Myh10 | Yes |
| NM_001106138 | 599 | 74 | 0 | 2.42E-149 | 245.7 | 1.11E-146 | Psmg2 | Yes |
| NM_001107494 | 199 | 71 | 0 | 3.59E-142 | 235.8 | 1.47E-139 | B9d2 | Yes |
| NM_001013189 | 599 | 67 | 0 | 1.08E-132 | 222.5 | 2.83E-130 | Pqlc1 | Yes |
| NM_001109515 | 399 | 63 | 0 | 2.56E-123 | 209.2 | 5.71E-121 | Slc25a28 | Yes |
| NM_001014185 | 599 | 63 | 0 | 2.56E-123 | 209.2 | 5.71E-121 | Srsf6 | Yes |
| NM_017049 | 599 | 61 | 0 | 1.14E-118 | 202.6 | 2.20E-116 | *Slc4a3* | Yes |
| NM_001008283 | 599 | 60 | 0 | 2.35E-116 | 199.2 | 4.31E-114 | Mettl23 | Yes |
| NM_001100978 | 599 | 59 | 0 | 4.75E-114 | 195.9 | 7.94E-112 | Tmem127 | Yes |
| NM_031784 | 599 | 58 | 0 | 9.47E-112 | 192.6 | 1.48E-109 | Pias3 | Yes |
| NM_001034157 | 599 | 58 | 0 | 9.47E-112 | 192.6 | 1.48E-109 | Gpank1 | Yes |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NM_134353 | 399 | 57 | 0 | 1.86E-109 | 189.3 | 2.62E-107 | Pabpc1 | Yes |
| NM_024002 | 599 | 56 | 0 | 3.57E-107 | 186.0 | 4.78E-105 | Secisbp2 | Yes |
| NM_001008300 | 799 | 56 | 0 | 3.57E-107 | 186.0 | 4.78E-105 | Nedd4l | Yes |
| NM_001191102 | 399 | 55 | 0 | 6.76E-105 | 182.6 | 8.29E-103 | Phtf1 | Yes |
| NM_001113783 | 599 | 54 | 0 | 1.26E-102 | 179.3 | 1.40E-100 | Golt1b | Yes |
| NM_001106758 | 399 | 53 | 0 | 2.30E-100 | 176.0 | 2.41E-98 | Cinp | Yes |
| NM_001108729 | 399 | 52 | 0 | 4.12E-98 | 172.7 | 3.94E-96 | Pmel | Yes |
| NM_001013124 | 199 | 52 | 0 | 4.12E-98 | 172.7 | 3.94E-96 | Ung | Yes |
| NM_001005879 | 599 | 50 | 0 | 1.25E-93 | 166.0 | 1.05E-91 | Gfpt1 | Yes |
| NM_001025018 | 599 | 49 | 0 | 2.12E-91 | 162.7 | 1.61E-89 | Dram2 | Yes |
| NM_053797 | 199 | 49 | 0 | 2.12E-91 | 162.7 | 1.61E-89 | Crnkl1 | Yes |
| NM_001134532 | 599 | 49 | 0 | 2.12E-91 | 162.7 | 1.61E-89 | Qrich1 | Yes |
| NM_001109452 | 599 | 49 | 0 | 2.12E-91 | 162.7 | 1.61E-89 | Mplkip | Yes |
| NM_001008764 | 599 | 48 | 0 | 3.52E-89 | 159.4 | 2.36E-87 | Psenen | Yes |
| NM_001134553 | 599 | 48 | 0 | 3.52E-89 | 159.4 | 2.36E-87 | Ubn2 | Yes |
| NM_053766 | 599 | 48 | 0 | 3.52E-89 | 159.4 | 2.36E-87 | Wbp4 | Yes |
| NM_001007745 | 599 | 47 | 0 | 5.73E-87 | 156.1 | 3.17E-85 | Tmem43 | Yes |
| NM_001166676 | 399 | 47 | 0 | 5.73E-87 | 156.1 | 3.17E-85 | Parpbp | Yes |
| NM_001025756 | 599 | 47 | 0 | 5.73E-87 | 156.1 | 3.17E-85 | Tmem186 | Yes |
| NM_001007683 | 599 | 47 | 0 | 5.73E-87 | 156.1 | 3.17E-85 | Glt8d1 | Yes |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NM_001108029 | 399 | 46 | 0 | 9.14E-85 | 152.8 | 4.42E-83 | Arid4a | Yes |
| NM_001170539 | 799 | 46 | 0 | 9.14E-85 | 152.8 | 4.42E-83 | Stt3b | Yes |
| NM_001100679 | 199 | 46 | 0 | 9.14E-85 | 152.8 | 4.42E-83 | Kdm2b | Yes |
| NM_001006984 | 399 | 45 | 0 | 1.43E-82 | 149.4 | 6.21E-81 | Vezt | Yes |

**Sertoli cell specific H3K4me3 enriched regions.** To gain insight on potential regulatory regions specific to Sertoli cells, subtractive analysis identified enriched regions specific to the Sertoli cell list (i.e. regions not also present in the myoid cell list) for each genome build (unmasked, masked and the upstream promoter region). Enrichment was determined by analyzing the relative fold change between the number of sequence reads within the identified regions between Sertoli and myoid cell populations. The top 40 enriched locations (based on highest fold change) are listed in Tables 3-8, 3-9 and 3-10 for the unmasked, masked and 1000 bp upstream region genome assemblies.

Table 3-8 identifies the top 40 Sertoli cell specific enriched regions using the unmasked genome build for alignment. Of these, 12 are located within introns and 28 are intergenic. This differs significantly from the top regions identified in Sertoli cells as the Sertoli cell specific regions correspond to introns instead of 5' flanking/exon1 regions and more intergenic regions were identified. Table 3-9 identifies the top 40 Sertoli cell specific peaks identified using the masked genome build for alignment. Of these, 21 are associated with introns/exons and 19 are intergenic. Again, this differs from the Sertoli cell list in that identified regions correspond to introns instead of 5' flanking/exon1 regions and an increase in the number of intergenic regions. The number of reads used to identify these regions is lower than the number identified in the Sertoli cell lists. This likely because many of the 5' flanking and Exon1 regions were also identified in myoid cells thus removing them from this list.

Comparison of the two genome assemblies only identified four regions that corresponded to both lists. Two of the regions were intergenic and two were within gene introns. This is a smaller number of cross-identified regions when compared to the Sertoli and myoid cell lists.

Table 3-10 identifies peaks located 1000 bp upstream of known TSS and only two of these regions (Fshr and Tinagl1) were also found in the masked genome assembly top 40 list. Gene expression was determined by searching the GEO expression database identifying genes expressed in E13 Sertoli cells (indicated by * within the table) or by general testis expression. All genes analyzed except one (RT1-M3-1) were expressed in E15 mouse Sertoli cells. This indicates that H3K4me3 enrichment within the upstream promoter region can be associated with gene expression in Sertoli cells.

**Table 3-8. Top 40 Sertoli cell specific H3K4me3 enriched regions- Unmasked genome assembly.**

| Chromosome Position | Size (bp) | SC Reads | MC Reads | P-value SC vs MC | Fold Change SC vs MC | FDR SC vs MC | Gene Association | Location |
|---|---|---|---|---|---|---|---|---|
| chr6:22614600-22616799 | 2199 | 692 | 13 | 0 | 90.6 | 0 | *Lhcgr, Fshr* | Intergenic |
| chr4:212849200-212849599 | 399 | 26 | 0 | 4.46E-36 | 49.4 | 2.69E-33 | Ly49i9 | Intron 5 |
| chr7:78006000-78006599 | 599 | 18 | 0 | 5.06E-23 | 34.8 | 6.05E-21 | Rims2 | Intron 2 |
| chr16:242600-242999 | 399 | 18 | 0 | 5.06E-23 | 34.8 | 6.05E-21 | Rps24 | Intergenic |
| chr10:4446400-4446799 | 399 | 17 | 0 | 1.76E-21 | 32.9 | 1.63E-19 | Rpl39l, Grin2a | Intergenic |
| chr20:54586000-54586199 | 199 | 17 | 0 | 1.76E-21 | 32.9 | 1.63E-19 | Hace1, Grik2 | Intergenic |
| chr1:37420400-37420999 | 599 | 16 | 0 | 5.81E-20 | 31.1 | 4.16E-18 | Irx1, Med10 | Intergenic |
| chr10:35706600-35706999 | 399 | 16 | 0 | 5.81E-20 | 31.1 | 4.16E-18 | Rufy1, Adamts2 | Intergenic |
| chr1:249824400-249824799 | 399 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | Pip5k1b, Fam122a | Intergenic |
| chr2:59855400-59855599 | 199 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | Pde4d | Intron 2 |
| chr3:163447200-163447799 | 599 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | Chd6, Ptprt | Intergenic |
| chr5:168833800-168833999 | 199 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | Fbxo2, Ubiad1 | Intergenic |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr6:32633200-32633599 | 399 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | Alk | Intron 2 |
| chr7:26175800-26176399 | 599 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | LOC100910996 | Intron 2 |
| chr8:120024800-120024999 | 199 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | Arpp21, Pdcd6ip | Intergenic |
| chr10:75907200-75907799 | 599 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | RGD1311564, Dhx40 | Intergenic |
| chr14:75860400-75860799 | 399 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | Rab26, Hs3st1 | Intergenic |
| chr18:75687000-75687199 | 199 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | Slc14a2, Nfatc1 | Intergenic |
| chr19:47217400-47218399 | 999 | 15 | 0 | 1.81E-18 | 29.3 | 9.79E-17 | Nr3c2, Ces2g | Intergenic |
| chr2:44358400-44359399 | 999 | 30 | 1 | 5.25E-43 | 28.4 | 5.72E-40 | S100z, F2rl1 | Intergenic |
| chr2:115225600-115225999 | 399 | 14 | 0 | 5.31E-17 | 27.4 | 2.22E-15 | Stmn2, Il7 | Intergenic |
| chr2:263847400-263847799 | 399 | 14 | 0 | 5.31E-17 | 27.4 | 2.22E-15 | Stpg2, Pdha2 | Intergenic |
| chr6:38431400-38432399 | 999 | 14 | 0 | 5.31E-17 | 27.4 | 2.22E-15 | *Adcy3* | Intron 3 |
| chr6:156659600-156659999 | 399 | 14 | 0 | 5.31E-17 | 27.4 | 2.22E-15 | Macc1, Tmem196 | Intergenic |
| chr15:62627400-62628199 | 799 | 14 | 0 | 5.31E-17 | 27.4 | 2.22E-15 | Serp2, Dnajc15 | Intergenic |
| chr19:48299200-48299399 | 199 | 14 | 0 | 5.31E-17 | 27.4 | 2.22E-15 | Zdhhc1 | Intron 5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr6:38431400-38432399 | 999 | 14 | 0 | 5.31E-17 | 27.4 | 2.22E-15 | *Adcy3* | Intron 2 |
| chr1:264882000-264882399 | 399 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Plce1 | Intron 7 |
| chr5:160864600-160865199 | 599 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Vwa5b1, Ubxn10 | Intergenic |
| chrX:151098600-151098999 | 399 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Slitrk2 | Promoter, Exon 1 |
| chr8:122260400-122260799 | 399 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Cmtm8, Gpd1l | Intergenic |
| chrX:116011400-116011799 | 399 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Lima1, Lhfpl1 | Intergenic |
| chr2:107593000-107593599 | 599 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Zfp445, LOC499565 | Intergenic |
| chr10:97268800-97269199 | 399 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Rgs9 | Intron 7 |
| chr15:83190000-83190399 | 399 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Pcdh9, Klhl1 | Intergenic |
| chr10:69660400-69660999 | 599 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Tmem132e, Cct6b | Intergenic |
| chr7:47696600-47696999 | 399 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Tmtc2, Mettl25 | Intergenic |
| chr4:25720400-25720799 | 399 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | Cldn12, Cdk14 | Intergenic |
| chr16:3622000-3622599 | 599 | 13 | 0 | 1.46E-15 | 25.6 | 4.73E-14 | *Erc2* | Intron 6 |
| chr2:147693800-147694599 | 799 | 12 | 0 | 3.75E-14 | 23.8 | 9.24E-13 | *Pgrmc2, Phf17* | Intergenic |

**Table 3-9. Top 40 Sertoli cell specific H3K4me3 enriched regions - Masked genome assembly.**

| Chromosome Location | Size (bp) | SC Reads | MC Reads | P-value SC vs MC | Fold Change SC vs MC | FDR SC vs MC | Gene Association | Location |
|---|---|---|---|---|---|---|---|---|
| chr6:22614400-22616799 | 2399 | 622 | 13 | 0 | 53.0 | 0 | *Lhcgr, Fshr* | Intergenic |
| chr2:214882600-214884399 | 1799 | 14 | 0 | 2.56E-14 | 17.9 | 6.91E-12 | RGD1563667 | Intron 2 |
| chr5:152103400-152104599 | 1199 | 14 | 0 | 2.56E-14 | 17.9 | 6.91E-12 | *Tinagl1* | Intron 4 |
| chr6:112993400-112995799 | 2399 | 14 | 0 | 2.56E-14 | 17.9 | 6.91E-12 | Smoc1 | Intron 5 |
| chr9:49600000-49601999 | 1999 | 14 | 0 | 2.56E-14 | 17.9 | 6.91E-12 | Fhl2 | Intron 2 |
| chr2:147706800-147707799 | 999 | 13 | 0 | 4.58E-13 | 16.7 | 7.25E-11 | *Pgrmc2, Phf17* | Intergenic |
| chr7:121742400-121744399 | 1999 | 13 | 0 | 4.58E-13 | 16.7 | 7.25E-11 | Enthd1 | Intron 2 |
| chr8:47861600-47862799 | 1199 | 13 | 0 | 4.58E-13 | 16.7 | 7.25E-11 | Ube4a | Intron 13 |
| chr16:3622000-3622599 | 599 | 13 | 0 | 4.58E-13 | 16.7 | 7.25E-11 | *Erc2* | Intron 5 |
| chrX:115954400-115954999 | 599 | 13 | 0 | 4.58E-13 | 16.7 | 7.25E-11 | Lima1 | Intron 8 |
| chr8:118291600-118292799 | 1199 | 12 | 0 | 7.66E-12 | 15.5 | 7.84E-10 | Ccdc12 | Intron 1 |
| chr10:71375400-71376399 | 999 | 12 | 0 | 7.66E-12 | 15.5 | 7.84E-10 | Tada2a, Acaca | Intergenic |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr11:36867800-36869199 | 1399 | 12 | 0 | 7.66E-12 | 15.5 | 7.84E-10 | Runx1, RGD1562683 | Intergenic |
| chr11:60170600-60171799 | 1199 | 12 | 0 | 7.66E-12 | 15.5 | 7.84E-10 | Morc1, Cd96 | Intergenic |
| chr12:26846800-26848199 | 1399 | 12 | 0 | 7.66E-12 | 15.5 | 7.84E-10 | Wbscr27, Eln | Intergenic |
| chr1:18775000-18775599 | 599 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Ptprk | Intron 7 |
| chr1:90150600-90151399 | 799 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | RGD1565712 | Exon 11 |
| chr2:204677800-204679399 | 1599 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Fam160a1, Rps3a | Intergenic |
| chr3:15744400-15745399 | 999 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Ass1, Ncs1 | Intergenic |
| chr5:171596400-171597399 | 999 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Tnfrsf9 | Intron 1 |
| chr6:144552800-144554599 | 1799 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Exoc3l4, RGD1560608 | Intergenic |
| chr7:28894800-28895799 | 999 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Ccdc53 | Intron 5 |
| chr8:22324600-22325999 | 1399 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | S1pr5 | UTR |
| chr8:114397800-114398999 | 1199 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Poc1a | Intron 8 |
| chr10:40582000-40582999 | 999 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Sparc | Intron/Exon 4-5 |
| chr10:70243600-70244799 | 1199 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Ap2b1 | Intron 19 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr11:67106400-67106999 | 599 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Nr1i2, Cox17 | Intergenic |
| chr13:118288200-118290399 | 2199 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Cd34, Cd46 | Intergenic |
| chr14:86330600-86331799 | 1199 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Kremen1, Xbp1 | Intergenic |
| chrX:56324400-56324999 | 599 | 11 | 0 | 1.19E-10 | 14.3 | 8.11E-09 | Il1rapl1 | Intron 4 |
| chr1:172095400-172096399 | 999 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Relt, P2ry6 | Intergenic |
| chr1:273831400-273832799 | 1399 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Wbp1l | Intron 2, Exon 2 |
| chr4:225703000-225704799 | 1799 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Ntf3 | Intron 1 |
| chr5:62604800-62606599 | 1799 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Cntfr, Rpp25l | Intergenic |
| chr5:170468800-170469999 | 1199 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Slc25a33, Spsb1 | Intergenic |
| chr6:118295200-118295999 | 799 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Numb, Psen1 | Intergenic |
| chr9:16307200-16308599 | 1399 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Vegfa, Mrpl14 | Intergenic |
| chr10:17156000-17157599 | 1599 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Neurl1b, Stk10 | Intergenic |
| chr10:75176000-75177799 | 1799 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Rnf43, Mtmr4 | Intergenic |
| chr10:105937200-105937999 | 799 | 10 | 0 | 1.71E-09 | 13.1 | 7.68E-08 | Sept9 | Intron 3 |

**Table 3-10. Top 40 Sertoli cell specific H3K4me3 enriched regions – 1000 bp upstream genome assembly.**
* indicates expression in E15 Sertoli cells

| Locus ID (1000 bp upstream) | Size (bp) | SC Reads | MC Reads | P-value SC vs MC | Fold Change SC vs MC | FDR SC vs MC | Gene Association | GEO Testis Expression |
|---|---|---|---|---|---|---|---|---|
| NM_001014248 | 199 | 21 | 0 | 3.24E-23 | 24.6 | 2.64E-21 | Cacul1 | Yes* |
| NM_199237 | 799 | 142 | 12 | 1.49E-105 | 12.3 | 4.37E-102 | *Fshr* | Yes* |
| NM_001106773 | 199 | 26 | 2 | 1.08E-22 | 10.1 | 8.32E-21 | Bpifc | Yes* |
| NM_001191958 | 199 | 21 | 2 | 5.79E-17 | 8.2 | 3.14E-15 | Zfyve16 | Yes* |
| NM_001109359 | 399 | 107 | 14 | 1.09E-61 | 8.1 | 6.37E-59 | Tmem229b | Yes* |
| NM_001017380 | 199 | 12 | 1 | 1.64E-11 | 7.3 | 3.79E-10 | Cyld | Yes* |
| NM_031797 | 399 | 45 | 7 | 1.76E-24 | 6.4 | 1.78E-22 | Cd82 | Yes* |
| NM_001013893 | 399 | 43 | 7 | 9.36E-23 | 6.1 | 7.42E-21 | Actl9b | Yes* |
| NM_001115031 | 399 | 70 | 12 | 4.54E-34 | 6.1 | 1.02E-31 | Cer1 | Yes* |
| NM_017081 | 399 | 129 | 26 | 6.55E-53 | 5.4 | 2.74E-50 | Hsd11b2 | Yes* |
| NM_022854 | 399 | 94 | 19 | 3.67E-39 | 5.3 | 1.19E-36 | Fabp9 | Yes* |
| NM_139324 | 199 | 18 | 3 | 9.03E-11 | 5.3 | 1.82E-09 | Ehd4 | Yes* |
| NM_001105987 | 199 | 21 | 4 | 4.43E-11 | 4.9 | 9.41E-10 | Tatdn3 | Yes* |
| NM_053582 | 199 | 34 | 7 | 1.70E-15 | 4.9 | 7.78E-14 | *Tinagl1* | Yes* |
| NM_001170439 | 399 | 29 | 6 | 1.65E-13 | 4.8 | 5.09E-12 | Evc | Yes* |
| NM_001107129 | 199 | 59 | 13 | 1.12E-23 | 4.8 | 9.92E-22 | Cops6 | Yes* |
| NM_001108251 | 199 | 20 | 4 | 2.75E-10 | 4.7 | 4.90E-09 | Ubqln2 | Yes* |
| NM_001128065 | 199 | 20 | 4 | 2.75E-10 | 4.7 | 4.90E-09 | Rpl36a | Yes* |
| NM_001207007 | 199 | 28 | 6 | 9.28E-13 | 4.6 | 2.57E-11 | Gamt | Yes* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NM_001135029 | 199 | 23 | 5 | 9.15E-11 | 4.5 | 1.84E-09 | Tmigd1 | Yes* |
| NM_001107620 | 199 | 58 | 14 | 1.91E-21 | 4.4 | 1.33E-19 | Ubash3a | Yes* |
| NM_001004081 | 399 | 34 | 8 | 7.67E-14 | 4.3 | 2.52E-12 | Mpi | Yes* |
| NM_001108797 | 199 | 37 | 9 | 2.01E-14 | 4.2 | 7.32E-13 | Fam117b | Yes* |
| NM_133593 | 399 | 37 | 9 | 2.01E-14 | 4.2 | 7.32E-13 | Ap3m1 | Yes* |
| NM_001033705 | 199 | 40 | 10 | 5.08E-15 | 4.2 | 2.21E-13 | Zfp451 | Yes* |
| NM_053677 | 399 | 40 | 10 | 5.08E-15 | 4.2 | 2.21E-13 | Chek2 | Yes* |
| NM_053961 | 399 | 32 | 8 | 1.81E-12 | 4.1 | 4.86E-11 | Erp29 | Yes* |
| NM_019222 | 199 | 10 | 2 | 2.94E-06 | 4.1 | 2.03E-05 | Coro1b | Yes* |
| NM_001007740 | 199 | 28 | 7 | 3.44E-11 | 4.1 | 7.52E-10 | Vps26a | Yes* |
| NM_001039016 | 199 | 31 | 8 | 8.40E-12 | 4.0 | 1.99E-10 | Zdhhc9 | Yes* |
| NM_032071 | 199 | 20 | 5 | 1.28E-08 | 3.9 | 1.62E-07 | Synj2 | Yes* |
| NM_053366 | 199 | 20 | 5 | 1.28E-08 | 3.9 | 1.62E-07 | Rab6a | Yes* |
| NM_022921 | 199 | 20 | 5 | 1.28E-08 | 3.9 | 1.62E-07 | RT1-M3-1 | Yes |
| NM_171996 | 399 | 44 | 12 | 5.71E-15 | 3.9 | 2.42E-13 | Decr2 | Yes* |
| NM_021740 | 399 | 37 | 10 | 4.64E-13 | 3.9 | 1.36E-11 | Ptma | Yes* |
| NM_017325 | 399 | 30 | 8 | 3.79E-11 | 3.9 | 8.17E-10 | Runx1 | Yes* |
| NM_001107058 | 199 | 23 | 6 | 3.11E-09 | 3.8 | 4.57E-08 | Map3k3 | Yes* |
| NM_024359 | 999 | 86 | 25 | 3.12E-25 | 3.7 | 3.51E-23 | Hif1a | Yes* |
| NM_001047901 | 199 | 19 | 5 | 6.08E-08 | 3.7 | 6.12E-07 | Ankle2 | Yes* |
| NM_001107600 | 199 | 22 | 6 | 1.40E-08 | 3.7 | 1.74E-07 | Fbxw4 | Yes* |

## Discussion

The mapping of post-translationally modified histone tails has become a vast field of study and has led to the identification of specific histone modifications that are correlated with various states of sequence activity. Some histone marks like Histone 4 acetylation tend to mark larger regions of active chromatin while others like Histone 2 lysine 27 tri-methylation and H3K4me3 correspond to a more specific function and localize to smaller genomic regions like promoters and enhancers. For our studies, we took advantage of the latter and used H3K4me3 as a marker to identify locally active regions within P15 Sertoli cells. Six databases were generated using this mark, three corresponding to Sertoli cells and three corresponding to Sertoli cell specific enrichment using myoid cells to filter out any cross identified regions. Each database was generated using three genome builds: unmasked, masked and a build containing only the upstream promoter region of known genes.

We included the three different genome assemblies in our analysis. The unmasked and masked genomes were each used to identify genome wide association patterns associated with H3K4me3. The unmasked assembly contains repetitive DNA sequence and was included because these regions are relevant to genome function and transcriptional regulation (Kidwell and Lisch 2000; Waterston, Lindblad-Toh et al. 2002; Feschotte 2008). Although relevant, repetitive regions are often difficult to map as a single read may align to numerous repetitive sites within the genome and alignment programs are only capable of aligning a single read to a single location. This can cause over/under representation of

enrichment at various locations. The inclusion of repetitive sequences in alignments can also mask significant enrichment in non-repetitive sequences when enrichment levels are evaluated based on fold change between and input sample. For this reason, we also included a masked genome assembly that did not contain repetitive sequences. The third genome assembly that was included contained only genomic sequence corresponding to 1000 bp upstream of transcriptional start sites. We included this genome assembly to generate a list of enriched regions pertaining to the 5' flanking sequences of all known genes (~13,450).

For our analysis, we analyzed the H3K4me3 enrichment patterns of two cell types- Sertoli and peritubular myoid cells. Both cell types are found within the testis, but each play a very different role in testis function. Sertoli cells are found within the tubules and foster the environment throughout spermatogenesis. Around P15 the Sertoli cells have ceased dividing and are terminally differentiating, preparing for the initiation of spermatogenesis. Peritubular myoid cells are found on the outside of the tubule. The role of the myoid cell is largely to assist in the movement of fluid through the tubules. Myoid cells were included in this study as a control to allow for the identification of Sertoli cell specific enriched H3K4me3.

Initial Sertoli cell analysis identified 37, 371, 26, 804 and 2,372 enriched regions in the unmasked, masked and 1000 bp upstream genome assemblies respectively; and Myoid cell analysis identified 59,441, 27,509 and 2,015 enriched regions respectively. With respect to the unmasked genome, the myoid cell genome contained 31,932 more enriched regions than the Sertoli cell genome. This is likely explained by the differential differentiation states between the two cells. At P15

Sertoli cells have undergone differentiation and are no longer actively dividing. As such, they contain more heterochromatin that the actively dividing myoid cells. Previous studies have indicated that H3K4me3 enrichment is higher in euchromatic vs. heterochromatic regions (Li, Wang et al. 2008; Radman-Livaja, Liu et al. 2010; Yin, Sweeney et al. 2011).

Analysis of the 1000 bp 5' flanking regions identified 2,372 enriched regions in the Sertoli cells and 2,015 enriched regions in the myoid cells. Numerous studies have indicated that enrichment of H3K4me3 at promoters and across transcriptional start sites denotes transcriptional activation. Published expression data from the Griswold lab indicates that there are ~2,800 expressed genes in P20 rat Sertoli cells indicating that there may be a correlation between expressed genes and 5' enrichment of H3K4me3 within Sertoli and myoid cells as in other studies cell types (McLean, Friel et al. 2002).

To identify Sertoli cell specific H3K4me3 enrichment, the Sertoli cell and myoid cell datasets were compared and subtractive analysis identified 22,914, 9,777 and 944 enriched regions for the unmasked, masked and 1000 bp upstream genome assemblies respectively. Previous studies have identified roughly 700 Sertoli cell specific genes expressed at P22 implying that a majority of the regions identified in the unmasked and masked genome are likely intergenic (Ryser, Glauser et al. 2011). This also implies that some of the regions identified in the upstream promoter region as Sertoli cell specific must be associated with silent or weakly expressed genes.

For analysis of specific regions within our databases, the top 40 regions were selected based on the number of reads within the sample compared to a control sample. Within the Sertoli cell unmasked and masked genome assembly datasets, seven genes were cross-identified (Mfsd8, Hn1l, Ptar1, Rad51c, Zfp40, Nup205 and RGD1303003) only three have been studied in reproductive tissues. Hn1l abundant in testis and skeletal muscle during embryo development (Zhou, Wang et al. 2004), however its role within the testis has not been further explored. Zfp40 transcripts were identified in spermatogonia, spermatocytes, and in testis cords in the genital ridge as well as in oocytes and follicle cells in the ovary(Noce, Fujiwara et al. 1993), however the role of Zfp40 in Sertoli cells has not been studied. Rad51c is implicated in breast and ovarian cancer and is involved in homologous recombination(Smeenk, de Groot et al. 2010; Zheng, Zhang et al. 2010; Monsees, Kraft et al. 2011), but its role within Sertoli cells has not been explored.

Within the Sertoli cell specific database, the top candidate identified in both the masked and unmasked genomes corresponds to a region that lies ~98kb from the 5' end of *Lhcgr* and ~140kb from the 3' end of *Fshr*. Expression of both of these genes is known for Sertoli cells and only *Fshr* is expressed within this cell type. The upstream promoter region of *Fshr* was also identified within the Sertoli cell specific database as the second most enriched region when compared to myoid cells. Research indicates that the *Fshr* gene is likely regulated by an unidentified distal regulatory element(Hermann, Hornbaker et al. 2007), and the role of this region will be further explored in Chapter 4 of this text.

Expression data is not available at this time for P15 Sertoli cells so GEO expression databases were searched for gene expression profiles. Analysis of the top 40 Sertoli cell specific genes identified by H3K4me3 enrichment within the 5' flanking region indicates that all of the identified genes are expressed in E13 Sertoli cells. This implies that again that H3K4me3 enrichment within upstream promoter regions may correlate with gene expression in Sertoli cells.

It is important to note that while we often think of genomes as being completely sequenced this is often on the case. While this data provides important information regarding H3K4me3 enrichment across the *Rattus norvegi*cus genome, the current rn5 genome build is only ~90% complete. Currently there is no sequence data for the Y chromosome in the current build and there are numerous regions across other chromosomes that haven't been fully sequenced. Thus, ChIP-Seq data will need to be realigned and reanalyzed with each new release of the *Rattus norvegicus* genome.

Taken together these datasets provide a snapshot of H3K4me3 within P15 Sertoli cells and support the idea that functional roles of H3K4me3 identified in other cell types may apply to Sertoli cells. These databases also highlight numerous intergenic regions enriched with H3K4me3 that may lead to the identification of new regulatory regions important in Sertoli cell function. Future work combining the H3K4me3 enrichment profiles with a database highlighting gene expression will also provide a powerful tool useful in the identification of genes and regulatory regions important in Sertoli cell function and differentiation.

# Chapter 4

## A distal 3' *cis*-element regulates the Follicle Stimulating Hormone Receptor gene in *Rattus norvegicus* Sertoli cells.

## Abstract

The follicle stimulating hormone receptor (FSHR) is a G-protein coupled receptor, which is responsible for conveying signals form the pituitary glycoprotein FSH. its regulation is critical to physiologic response to FSH, and thus to gonad development and regulation of fertility. To date, transcriptional mechanisms responsible for its exquisite cell-specificity remain poorly understood. Prior studies have identified regulatory elements and their proteins that function through the promoter region, but have provided few clues as to the mechanism of cell-specific expression. Previous studies have identified that regions outside the *Fshr* promoter are responsible for driving cell-specific expression of *Fshr.* As knowledge of gene regulation progresses, key regulatory domains have been identified that lie outside the proximal promoter, sometimes hundreds of kilobases away. Only recently have technological advances provided sufficient tools to uncover these distal regulatory elements. To help uncover distal regulatory elements that direct expression of *Fshr*, we used a combination of bioinformatic and molecular assays. Comparative genomics identified thirty-one ECRs surrounding *Fshr* and histone modification analysis identified a single site, ECR 1f, indicative of Sertoli cell specific activity. Transient transfection analysis of ECR 1f in rat Sertoli and myoid cells demonstrate that this region represses the *Fshr* promoter *in vitro*. Block deletion studies of ECR1f identified a specific region of that, when deleted, eliminated the repressive action of ECR1f on the *Fshr* promoter *in vitro*. EMSA analysis of this region identified two protein-binding sites, one site bound Gata4 and another bound an unidentified protein *in vitro.* Collectively, these studies have identified a potential *cis*-regulatory

element capable of repressing the *Fshr* promoter *in vitro.* Future studies will investigate the role these proteins play in transcriptional regulation of *Fshr in vivo.*

# Introduction

Accurate signaling between the pituitary and the gonads is critical for reproductive function and requires timely release of signaling molecules from the pituitary and proper spatio-temporal expression of their receptors on the target cells. Within the *hypothalamic–pituitary–gonadal axis*, FSH is an important glycoprotein hormone produced by the pituitary that specifically targets the gonads by binding its receptor, follicle stimulating hormone receptor (FSHR), which is expressed only on testicular Sertoli cells and ovarian granulosa cells. FSHR is a G protein-coupled receptor and the binding of FSH induces signaling cascades that affect many cellular and physiological processes including cell division, differentiation, follicle development, hormone secretion and fertility (Tindall, Schrader et al. 1974; Steinberger, Heindel et al. 1975; Chemes, Dym et al. 1979; Handel and Eppig 1979). Thus, proper reproductive function requires accurate expression of *Fshr.*

In males, the physiologic role of FSHR and FSH differs with the developmental sage of the testis. During early testis development (E13-P15), FSH induces proliferation of immature Sertoli cells, while later, around P15, the action of FSHR action on Sertoli cells changes from supporting Sertoli cell proliferation to supporting their differentiation (Murphy 1965; Dorrington, Roller et al. 1974; Means and Huckins 1974; Orth 1984; Chaudhary, Whaley et al. 1996). The period of growth and the timing of differentiation are critical as the Sertoli cells cease dividing prior to puberty and the final number of Sertoli cells sets the upper limit of sperm

production by the testes (Steinberger and Steinberger 1971). As Sertoli cell proliferation ceases, proper differentiation ensures development of mature Sertoli cells, which are critical for supporting spermatogenesis (Orth, Gunsalus et al. 1988).

In the female, the primary role of FSH is to support folliculogenesis. Similar to the testis, FSH targets a specific cell type within the ovary, the granulosa cells, and supports the proliferation of Granulosa cells, a critical step in folliculogenesis. FSH also plays a role in regulating the synthesis of aromatase, the key enzyme involved in the conversion of testosterone to estradiol (Gore-Langton and Dorrington 1981). Estradiol is the bioactive form of estrogen responsible for stimulating endometrial growth prior to ovulation, a critical event for successful embryo implantation. Defects in FSH signaling, either due to a pituitary or receptor issue, result in complete infertility in the female, again highlighting the importance of this signaling pathway and the proper expression of *Fshr.*

The goal of this project is to elucidate the transcriptional mechanisms involved in cell specific activation of the follicle stimulating hormone receptor (FSHR) gene. As discussed above, regulation of FSHR at a cellular level plays a critical role in mediating the physiologic responses to FSH and thus correct spatio-temporal expression of *Fshr,* which requires specific regulatory elements within the *Fshr* locus, is critical for proper cellular function. We are focusing our studies on *Fshr* regulation in P15 Sertoli cells, and we specifically want to identify the distal regulatory region involved in correct spatio-temporal expression of *Fshr.*

Since 1992, when Fshr was first cloned, numerous studies have sought to determine the transcriptional mechanisms behind its exquisite cell-specificity

(Heckert and Griswold 1991; Heckert, Daley et al. 1992; Gromoll, Dankbar et al. 1994; Linder, Heckert et al. 1994; Findlay and Drummond 1999; Heckert, Sawadogo et al. 2000; Heckert 2001; Xing and Sairam 2001; Xing and Sairam 2002; Hermann and Heckert 2005; Wunsch, Ahda et al. 2005; Hermann and Heckert 2007; Zaidi, Zhu et al. 2007; Lu, Yang et al. 2009; Perez-Solis, Macias et al. 2009). To date, studies have identified several regulatory elements, and their cognate binding factors, within the *Fshr* promoter region, but have yielded few clues about how cell-specific expression is achieved (Heckert, Sawadogo et al. 2000; Heckert 2001; Kim and Griswold 2001; Levallet, Koskimies et al. 2001). Research has shown that the elements identified within the promoter (-5000 bp) control basal transcription but cannot direct correct spatio-temporal expression of *Fshr* (Linder, Heckert et al. 1994). Transgenic animals carrying a 5000 bp promoter/reporter construct or a 413 kb yeast artificial chromosome (YAC) containing the entire rat *Fshr* gene also failed to restrict gene expression to Sertoli cells (Hermann, Hornbaker et al. 2007). Thus, our knowledge of Fshr transcription combined with more recent insight on the role of distal elements suggests correct spatio-temporal expression is controlled by distal elements.

Current knowledge of *cis*-acting regulatory elements shows they can lie anywhere from a few hundred base pairs to 1 Mb away from the transcriptional start site and function as enhancers, silencers and locus control regions (Amouyal 1991; Ruh, Dunn et al. 1996; Ringrose, Chabanis et al. 1999; Engel and Tanimoto 2000; Liu and Garrard 2005; Wicks and Knight 2011; Guo, Monahan et al. 2012). Physical interaction between the promoter and the *cis*-regulatory regions are

observed, indicating that chromatin looping occurs. Activation by *cis*-regulatory elements can occur by several mechanisms. Androgen receptor (AR) activation of the Prostate specific antigen (PSA) enhancer element increases transcription by recruiting RNA Polymerase II (RNA PolII) to the *cis*-element, which then scans along the DNA until the promoter region is reached and activated (Wang, Carroll et al. 2005). Within the beta-globin locus, interactions between GATA-1 and FOG-1 mediate the interaction between the locus control region (*cis*-regulatory region) and the beta-globin promoter (Vakoc, Letting et al. 2005). In all, there are growing list of genes regulated by *cis*-elements, and the mechanisms that regulate the long-range interactions are now being identified.

To identify potential *cis*-regulatory elements important for cellular specificity, the present study evaluated *Fshr* and its intergenic sequences for regions of conservation and the presence of H3K4me3 unique to Sertoli cells, using myoid cells for comparison. These two techniques led to the identification of one conserved region that was tested for function using *in vitro* transcription and DNA binding assays. *In vitro* analysis identified a region repressive region within this *cis*-element that was capable of binding GATA-4 *in vitro.* Data derived from the current studies adds to our understanding of the transcriptional mechanisms required for *Fshr* expression, and how the receptor functions on a cellular level to regulate gamete development and maintain fertility.

## Materials and Methods

**Experimental Animals.** All experiments using animals were approved by the Institutional Animal Care and Use Committee of the University of Kansas Medical Center (Protocol 2009-1850), and performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

**Sertoli Cell and Myoid Cell Preparations.** Sertoli and myoid cells were prepared as previously described with slight modification (Steinberger, Heindel et al. 1975; Karl and Griswold 1990). Tubules were isolated from P15 Sprague Dawley rats and processed through a series of digestions releasing the myoid cells which, following centrifugation, remain in the supernatant while Sertoli cells and germ cells are collected in the pellet. During the isolation, myoid cells were removed from the tubules prior to complete digestion and subsequent release of the Sertoli and germ cells. The majority of germ cells remaining in the culture after digestion are spermatogonia as more advanced stages of germ cells are removed during tubule digestion. Freshly isolated Sertoli cells were used fresh for ChIP assays and contain a subpopulation of spermatogonia. For transfection analysis, Sertoli cells were and cultured for two days, shocked with 10 mM Tris pH 7.4 for 2-4 minutes to remove germ cells, and transfected. Myoid cells were cultured and passaged once prior to transfection. All cell cultures were incubated at 37°C in 5% $CO_2$.

**ChIP.** Chromatin was prepared as previously described with slight modifications (Aparicio, Geisberg et al. 2005). Prior to cross linking, Sertoli and myoid cells were

counted to establish cell numbers for micrococcal nuclease digestion. To cross-link chromatin, primary rat Sertoli and myoid cells were and incubated in 30 mL cell media containing 1% Formaldehyde for 10 minutes at room temperature. Cross-linking was stopped by the addition of 2.5 M glycine to a final concentration of 125mM. Cells were pelleted and washed two times in 10 mL ice-cold PBS containing freshly prepared Protease Inhibitors (Roche, South San Francisco, CA). Cell pellets were then washed three times in 20 mL ice-cold MC lysis buffer (10mM Tris-HCl pH 7.5, 10mM NaCl, 3mM $MgCl_2$, 0.5% NP-40) and incubated on ice for 10 minutes during the third wash. After the final wash was complete the cell pellets were vortexed (setting 3) for three minutes. The cells were pelleted, the supernatant removed and the pellet frozen by immersion in liquid nitrogen. The pellets were then thawed in an ice-water bath, resuspended to 4 x $10^7$ cells/mL in MNase buffer (10mM Tris-HCl pH 7.5, 10mM NaCl, 3mM $MgCl_2$, 1mM $CaCl_2$, 4% NP-40) and disrupted using 10 strokes with a Type B pestle in an ice-cold Dounce homogenizer to release chromatin. Chromatin suspensions were diluted with cold MNase buffer to a final concentration of 2 x $10^7$ cells/mL, Micrococcal nuclease (New England BioLabs M0247S) was added (120U per 2 x $10^6$ cells) and extracts were incubated for 30 minutes at 37°C. Nuclease digestion was stopped by the addition of (final concentration): 6mM EGTA, 1x protease inhibitors (Roche, South San Francisco, CA), 10% SDS and 200mM NaCl. Chromatin samples were aliquoted to 1 x $10^7$ cells (~500 μl) and stored at -80°C. Rat Sertoli and myoid chromatin was used for ChIP at an equivalent of 1 x $10^7$ cells per immunoprecipitation sample. Chromatin samples were thawed and cleared with 75 μl of a 50% slurry of Salmon sperm-

blocked protein A-agarose beads (Millipore 16-157) for one hour at 4°C with rotation. Agarose beads were then pelleted by centrifugation for one minute at 3900 x g at 4°C and the supernatants transferred to new tubes. The following antibodies were purchased and used for immunoprecipitation: Anti-acetyl-Histone H4 (Lys5, 8, 12, 16) (Millipore 06-866); Anti-trimethyl-Histone H3 (Lys4) (Millipore 07-473); Anti-trimethyl-Histone H3 (Lys27) (Millipore 07-449), IgG (Santa Cruz SC-2027), Anti-dimethyl-Histone H3 (Lys9) (Millipore 07-441) and Anti-acetyl-Histone H3 (Millipore 06-599). Five micrograms of primary antibody was added to the supernatants and the samples incubated at 4°C for 18 hours. Complex were precipitated with 100 μl of protein A-agarose slurry and incubated for one hour at 4°C with rotation. Chromatin and agarose beads were then transferred to Spin-X tubes (Corning, Tewksbury MA-#8160) for washing. Each wash (600 μl) was preformed for five minutes at RT on a rotator, followed by a one-minute spin at 300 x g. Wash conditions were as follows: 1 x Low Salt Wash (20mM Tris pH 8.1, 150mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS, protease inhibitors), 1 x High Salt Wash (20mM Tris pH 8.1, 500mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS, protease inhibitors), 1 x LiCl Wash (10mM Tris pH 8.1, 250mM LiCl, 1% NP-40, 1% Na Deoxycholate, 1mM EDTA, protease inhibitor), 2 x TE wash (10mM Tris pH 8.0, 1mM EDTA, protease inhibitor). Chromatin elution was achieved by incubating the agarose beads in 200 μl Elution buffer (100mM $NaHCO_3$, 1% SDS) for 15 minutes at RT with rotation. Supernatant was collected and NaCl added to a final concentration of 200mM. Samples were then incubated at 65°C for four hours to reverse the cross-links. Twenty micrograms of Proteinase-K was added and the

buffer concentration increased to 50mM Tris and 10mM EDTA. Samples were incubated for one hour at 45°C. DNA was isolated using the ChIP DNA Clean and Concentrator Kit (Zymo Research D5205) and eluted in 50 μl TE. Prior to Illumina sequencing, each ChIP-DNA sample was first analyzed to assess quality via qPCR using an Applied Biosystems 7900HT Fast Real-Time PCR System. Each reaction was preformed in triplicate and contained one microliter ChIP DNA, five microliters SyberGreen Master Mix (Applied Biosystems), and 300 nM forward and reverse primer. Primers for the *Fshr* promoter are FHSR Prom F 5'-TTTACTTGCCTGGAAGCGACTAA-3' AND FSHR Prom R 5'-GCTTGGAGAACGGGCAAA-3' and for the *Nrxn1* promoter are Nrxn1 Prom F 5'-GGGAGGAATCTGATCACTGTACTGT-3' AND Nrnx1 Prom R 5'-CATGATCTTCCCAATTGTCCAA-3'. The standard cycle was used, without any changes to the default parameters. A dissociation curve was run following the final cycle to determine if contamination was present. The SDS Software package 2.4 (Applied Biosystems) was used to calculate the cycle threshold (Ct) value for each primer set, which were adjusted such that the threshold was set at the initial phase of linear PCR amplification. Data was exported to a Microsoft Excel spreadsheet and analyzed as a percent of the input sample [100*2^(Adjusted input - Ct sample)]. Adjusted input was calculated based on the Ct value of the 1% input sample. qPCR was also used to verify regions identified by next generation sequencing as listed above.

**ChIP-DNA Sequencing.** ChIP DNA samples from two independent samples were pooled, analyzed via qPCR, dried and shipped to Cofactor Genomics LLC (St. Louis, MO) for library preparation and sequencing. Samples were prepared following Illumina library preparation protocols by Cofactor and sequenced on the Illumina GAIIx. The NEBNext® ChIP-Seq Library Prep Reagent Set for Illumina® (New England Biolabs), was used by Cofactor to prepare all samples for sequencing. Briefly, 10 ng ChIP DNA was treated with klenow and T4 DNA Polymerase to repair fragmented ends, cleaned and ligated to the NEBNext Adaptor sequence. The DNA was then cleaned and PCR amplified for 15 cycles using the NEBNext High-Fidelity 2X PCR Master Mix and Universal PCR Primers per the manufacturers protocol. The resultant library was cleaned using AMPure XP Beads, diluted and analyzed on the Agilent Bioanalyzer. Twenty picomoles of Library DNA were prepared for sequencing using the Illumina Cluster station. This ligated a llumina-supplied adapter sequence to one end of the DNA fragment, amplified the DNA clusters and hybridized the DNA to the flow-cell. The flow-cell was then loaded onto the Genome Analyzer IIx. Sequencing data output was in FASTQ files.

**Processing and Analysis of Sequencing Data.** ChIP-Seq data was analyzed through the K-INBRE Bioinformatics Core, University of Kansas, by Aaron Smalter Hall, Ph.D. FASTQ files representing eight lanes of single-end Illumina ChIP-Seq sequence tags produced by Cofactor Genomics were analyzed. FastQC (Babraham Bioinformatics-http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to test the quality of the sequence data for each sample prior to alignment.

FastQC is a quality control application for high throughput sequence data that provides an interactive application to review the results of several different QC checks relating to the quality of sequence tags. The Trim Galore! (Babraham Bioinformatics -http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) tool was used to trim low-quality base pairs, the Illumina primer and known adapter sequences using the following parameters: Quality Phred score cutoff: 51, Quality encoding type selected: ASCII+64, Adapter sequence: 'GATCGGAAGAGCGGTTCAGCAGGAATGCCG', Maximum trimming error rate: 0.1 (default), Minimum required adapter overlap (stringency): 1 bp and Minimum required sequence length before a sequence gets removed: 20 bp. The following command line was used: -f fastq -e 0.1 -q 51 -O 1 -a GATCGGAAGAGCGGTTCAGCAGGAATGCCG 'file name'.fq. Removal of duplicate sequence tag was performed using the "rmdup" module in the Samtools package (Sourceforge.net- http://samtools.sourceforge.net/). The BWA alignment tool (Sourceforge.net- http://bio-bwa.sourceforge.net/) was used to align the resulting files to the latest build of the *Rattus norvegicus* genome (rn5, May 2012) using the default parameters. Alignment was performed against two base library builds – a 'Soft Masked' and a 'Hard Masked'. The soft masked file contains the entire rn5 genome with repeat region denoted by lower case font while the hard masked genome masks all repeat regions to N's. These will be referred to as Unmasked and Masked respectively. Visualization of the mapped sequence tags was achieved using the UCSC Genome Browser using the BED, BEDGraph and BigWig file formats.

**UCSC Genome Browser Custom Track Settings.** To visualize the aligned sequence tags, bigwig files were imported into the UCSC Genome Browser using the 'add custom track' feature. The bigWig format is for display of dense, continuous data that will be displayed in the Genome Browser as a graph. Each track was imported based on the alignment of the sequence tag, i.e. positive or negative strand alignment. The following import command was used: track type=bigWig name="file name" description="" bigDataUrl=http://[ftp location of file]. Visualization settings were set as follows- Display mode: Full, Type of graph: bar, Track height: 40 pixels, Vertical viewing range: min-1 max-30, Data view scaling: use vertical viewing range setting, Always include zero: off, Transform function: Transform data point by: None, Window function: maximum, Smoothing window: off, Dray y indicator lines: at y=0.0: off at y=0 off.

**Transfection Vectors.** Vector inserts for ECR 1f, ECR 1f-a and 1f-b and ECR 1f Block Deletions were PCR amplified (primers listed in Table 4-1) from rat genomic DNA using iProof polymerase and PCR (98 C denaturing temperature, annealing 66 C, 30' x 35 cycles, and 72 C extension). The ECR 1f a/b clone was generated by restriction digest and ligation of the ECR 1f-a and ECR 1f-b fragments using a BamHI. Vectors for the Footprint mutations were amplified via Fusion PCR. Briefly, 5' and 3' PCR fragments were fused by PCR (98 C denature, 60 C anneal for 1 minute, 72 C extension for 3 minutes – for 3 cycles), ECR 1f-SacI F and ECR 1f-SacI R primers were added and the templates amplified (98 C denaturing temperature, annealing 66 C, 30' x 35 cycles, and 72 C extension). All digests used Fermentas Fast Digest

restriction enzymes (SacI, KpnI, BamHI and NdeI) and cloned into a pGL3-Basic vector, a 3:1 insert to vector ratio, containing either a -220/+123 or -2700/+123 bp fragment of the FSHR promoter in both 5' (SacI and KpnI) and 3' orientation (BamHI) (Figure 4-1) (Heckert, Daggett et al. 1998). 400,000 units of T4 DNA ligase (New England Biological) were used to ligate the insert and vector. Constructs were transformed to DH5 alpha competent cells, spread on LB+Ampicillin culture plates and allowed to grow overnight at 37°C. Plasmids used for transient transfections were prepared using the GeneJET Plasmid Miniprep kit (Fermentas, Maryland, MD), and confirmed by restriction nuclease digestion and DNA sequencing.

**Statistical Analysis.** Statistical analysis was performed using GraphPad Prism 6.0. For transient transfection analysis, ANOVA was used to determine statistical significance using the base vector (-220/+123 or -2700/+123) as the control. Sertoli and Myoid cell transfections were analyzed using two-way ANOVA and MSC1 cell transfections were analyzed using one-way ANOVA. ChIP-qPCR data was analyzed using one-way ANOVA, statistical significance was determined using the IgG sample as the control. For all analyses, statistical significance is marked on the graphs using asterisks as follows, * = p-value≤0.05, ** = p-value≤0.01, *** = p-value≤0.001 and **** = p-value≤0.0001

Table 4-1. Oligodeoxynucleotide primers used to generate ECR 1f transfection inserts.

| Primer | Seuqence | Region Amplified | Restriction Site (Vector insertion/internal) |
|---|---|---|---|
| rFSHR ECR 1f-SacI F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f | SacI/-- |
| rFSHR ECR 1f-BamHI F<br>rFSHR ECR 1f-BamHI R | 5'-CGCGGGATCCGAGCCTGTTTCTTGGATGCC-3'<br>5'-GGCCGGATCCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f | BamHI/-- |
| rFSHR ECR 1f-KpnI F<br>rFSHR ECR 1f-KpnI R | 5'-GGCCATGCGGGTAACGAGCCTGTTTCTTGGATGCC-3'<br>5'-CGCGGTATCGGTACCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f | KpnI/-- |
| rFSHR ECR 1f-a BamHI R<br>rFSHR ECR 1f-a BamHI F | 5'-GCTGGAGGATCCCACGACCCTGCCA-3'<br>5'-GCTCAGGATCCGTCTAAGCCCGTGTG-3' | ECR 1f-a | BamHI/-- |
| rFSHR ECR 1f-a KpnI R<br>rFSHR ECR 1f-a KpnI F | 5'-GCTGGAGGTACCCACGACCCTGCCA-3'<br>5'-CGTCAGGTACCGTCTAAGCCCGTGTG-3' | ECR 1f-a | KpnI/-- |
| rFSHR ECR 1f-b KpnI F<br>rFSHR ECR 1f-b KpnI R | 5'-GACGTGGTACCGGTTCCAGGTAAGAA-3'<br>5'-GTCGTGGTACCGAAGTAAAAATACTC-3' | ECR 1f-b | KpnI/-- |
| rFSHR ECR 1f-b BamHI F<br>rFSHR ECR 1f-b BamHI R | 5'-GACGTGGATCCGGTTCCAGGTAAGAA-3'<br>5'-GTCGTGGATCCGAAGTAAAAATACTC-3' | ECR 1f-b | BamHI/-- |
| rFSHR ECR 1f-SacI F<br>rFSHR ECR 1f-a BamHI R<br>rFSHR ECR 1f-b BamHI F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-GCTGGAGGATCCCACGACCCTGCCA-3'<br>5'-GACGTGGATCCGGTTCCAGGTAAGAA-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f a/b | SacI/BamHI |
| rFSHR ECR 1f Mut 1 SacI F<br>rFSHR ECR 1f-SacI R | 5'-CAGCCGATAGAGCTCCTCCATTCTTCCTGCAAAGG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 1 | SacI/-- |
| rFSHR ECR 1f-SacI F<br>rFSHR ECR 1f Mut 1 NdeI R<br>rFSHR ECR 1f Mut 2 NdeI F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGCCTTTGCAGGAAGAATGGAG-3'<br>5'-GCGCATGGCCATATGCATCCTGACAATATCCATTTTT-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 2 | SacI/NdeI |
| rFSHR ECR 1f-SacI F<br>rFSHR ECR 1f Mut 2 NdeI R<br>rFSHR ECR 1f Mt 3 NdeI F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGAAAAATGGATATTGTCAGGATG-3'<br>5'-GCGCATGGCCATATGTGACCAGCTACTGCTTGTGG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 3 | SacI/NdeI |

| | | | |
|---|---|---|---|
| rFSHR ECR 1f Mut 3 SacI F<br>rFSHR ECR 1f-SacI R | 5'-GCGCATGGCGAGCTCTGACCAGCTACTGCTTGTGG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 4 | SacI/-- |
| rFSHR ECR 1f-SacI F<br>rFSHR ECR 1f Mut 4 NdeI R<br>rFSHR ECR 1f Mut 5 NdeI F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGGAGTCTGAGGCCCTGAAAGA-3'<br>5'-GCGCATGGCCATATGTTGTGGATTGATTATTGAGTT-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 5 | SacI/NdeI |
| rFSHR ECR 1f-SacI F<br>rFSHR ECR 1f Mut 5 NdeI R<br>rFSHR ECR 1f Mut 6 NdeI F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGAACTCAATAATCAATCCACAA-3'<br>5'-GCGCATGGCCATATGAAAGGTTCACATTAATCTACAGTGG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 6 | SacI/NdeI |
| rFSHR ECR 1f-SacI F<br>rFSHR ECR 1f Mut 4 SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CGCGGTATCGAGCTCGAGTCTGAGGCCCTGAAAGA-3' | ECR 1f Block Deletion 7 | SacI/-- |
| rFSHR ECR 1f-SacI F<br>rFSHR ECR 1f Mut 2 NdeI R<br>1f BD 3.1 NdeI-F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGAAAAATGGATATTGTCAGGATG-3'<br>5'-GCGCATGGCCATATGCACAATCCAAAACTATGTTCC<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 3.1 | SacI/NdeI |
| rFSHR ECR 1f-SacI F<br>1f BD 3.2 NdeI-R<br>1f BD 3.2 NdeI-F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGGTTTCTCCTATCAGATTTAAACAAC-3'<br>5'-CCAGTCGAGCATATGCAGAAGGCATTCCTGTGATGAG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 3.2 | SacI/NdeI |
| rFSHR ECR 1f-SacI F<br>1f BD 3.3 NdeI-R<br>1f BD 3.3 NdeI-F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGTGGTTTTGCTTCTGAAAAAGG-3'<br>5'-CCAGTCGAGCATATGGGCGTTTAGTAATTTCAGTGG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 3.3 | SacI/NdeI |
| rFSHR ECR 1f-SacI F<br>1f BD 3.4 NdeI-R<br>1f BD 3.4 NdeI-F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGTTATGAATAATGTCTGAACTC-3'<br>5'-CCAGTCGAGCATATGAGCTACTGCTTGTGGTTTGAC-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 3.4 | SacI/NdeI |
| rFSHR ECR 1f-SacI F<br>rFSHR ECR 1f Mut 2 NdeI R<br>1f BD 3.4 NdeI-F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGAAAAATGGATATTGTCAGGATG-3'<br>5'-CCAGTCGAGCATATGAGCTACTGCTTGTGGTTTGAC-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 3.5 | SacI/NdeI |

| | | | |
|---|---|---|---|
| rFSHR ECR 1f-SacI F<br>1f BD 3.2 NdeI-R<br>1f BD 3.4 NdeI-F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGGTTTCTCCTATCAGATTTAAACAAC-3'<br>5'-CCAGTCGAGCATATGAGCTACTGCTTGTGGTTTGAC-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 3.6 | SacI/NdeI |
| rFSHR ECR 1f-SacI F<br>1f BD 3.3 NdeI-R<br>1f BD 3.4 NdeI-F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGTGGTTTTGCTTCTGAAAAAGG-3'<br>5'-CCAGTCGAGCATATGAGCTACTGCTTGTGGTTTGAC-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 3.7 | SacI/NdeI |
| rFSHR ECR 1f-SacI F<br>1f BD 3.2 NdeI-R<br>1f BD 3.3 NdeI-F<br>rFSHR ECR 1f-SacI R | 5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3'<br>5'-CCAGTCGAGCATATGGTTTCTCCTATCAGATTTAAACAAC-3'<br>5'-CCAGTCGAGCATATGGGCGTTTAGTAATTTCAGTGG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3' | ECR 1f Block Deletion 3.8 | SacI/NdeI |
| ECR1f mutFP 21 F<br>rFSHR ECR 1f-SacI R<br>ECR1f mutFP 21 R<br>rFSHR ECR 1f-SacI F | 5'-TCAGAACTGAGTTGTTTAAATCTGATAGGAG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3'<br>5'-CTCCTATCAGATTTAAACAACTCAGTTCTGA-3'<br>5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3' | ECR 1f mutated FP 21 | SacI |
| ECR1f mutFP 22 F<br>rFSHR ECR 1f-SacI R<br>ECR1f mutFP 22 R<br>rFSHR ECR 1f-SacI F | 5'-TCAGAAACTAACGCTTTAAATCTGATAGGAG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3'<br>5'-CTCCTATCAGATTTAAAGCGTTAGTTTCTGA-3'<br>5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3' | ECR 1f mutated FP 22 | SacI |
| ECR1f mutFP 23 F<br>rFSHR ECR 1f-SacI R<br>ECR1f mutFP 23 R<br>rFSHR ECR 1f-SacI F | 5'-TCAGAAACTAGTTGTGCTGCAGCTGTAGGAG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3'<br>5'-CTCCTACAGCTGCAGCACAACTAGTTTCTGA-3'<br>5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3' | ECR 1f mutated FP 23 | SacI |
| ECR1f mutFP 24 F<br>rFSHR ECR 1f-SacI R<br>ECR1f mutFP 24 R<br>rFSHR ECR 1f-SacI F | 5'-AGGTCTGTGTAATCAGTTCCTTTTTCAGAAGCAAAACCAC-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3'<br>5'-AACTGATTACACAGACCTTATCAGATTTAAACAACTAGT-3'<br>5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3' | ECR 1f mutated FP 24 | SacI |
| ECR1f mutFP 25 F<br>rFSHR ECR 1f-SacI R<br>ECR1f mutFP 25 R<br>rFSHR ECR 1f-SacI F | 5'-GTAGTCGTTTTTCAGAAGCAAAACCACAGAAGGC-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3'<br>5'-CGACTACAGTTTTGGATTGTGGTTTCTCC-3'<br>5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3' | ECR 1f mutated FP 25 | SacI |

| | | | |
|---|---|---|---|
| ECR1f FP 26mut- F<br>rFSHR ECR 1f-SacI R<br>ECR1f FP 26mut- R<br>rFSHR ECR 1f-SacI F | 5'-TCAGAAGCAATGATTACGAAGGCATCC-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3'<br>5'-GGATGCCTTCGTAATCATTGCTTCTGA-3'<br>5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3' | ECR 1f mutated FP 26 | SacI |
| ECR1f FP 27mut- F<br>rFSHR ECR 1f-SacI R<br>ECR1f FP 27mut- R<br>rFSHR ECR 1f-SacI F | 5'-GCATTCCTGTAGGTCTTTCAGACATT-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3'<br>5'-AATGTCTGAAAGACCTACAGGAATGC-3'<br>5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3' | ECR 1f mutated FP 27 | SacI |
| ECR1f FP 28mut-F<br>rFSHR ECR 1f-SacI R<br>ECR1f FP 28mut-R<br>rFSHR ECR 1f-SacI F | 5'-AGACATTATTGTGTTACTGTTTAGTAA-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3'<br>5'-TTACTAAACAGTAACACAATAATGTCT-3'<br>5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3' | ECR 1f mutated FP 28 | SacI |
| ECR1f FP 29mut-F<br>rFSHR ECR 1f-SacI R<br>ECR1f FP 29mut-R<br>rFSHR ECR 1f-SacI F | 5'-TAGTAATTTCTCACTGTAAGACGTGTCTCTG-3'<br>5'-CGCGGTATCGAGCTCCCTTCCTTCACTGTCTGCCTC-3'<br>5'-CAGAGACACGTCTTACAGTGAGAAATTACTA-3'<br>5'-GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC-3' | ECR 1f mutated FP 29 | SacI |

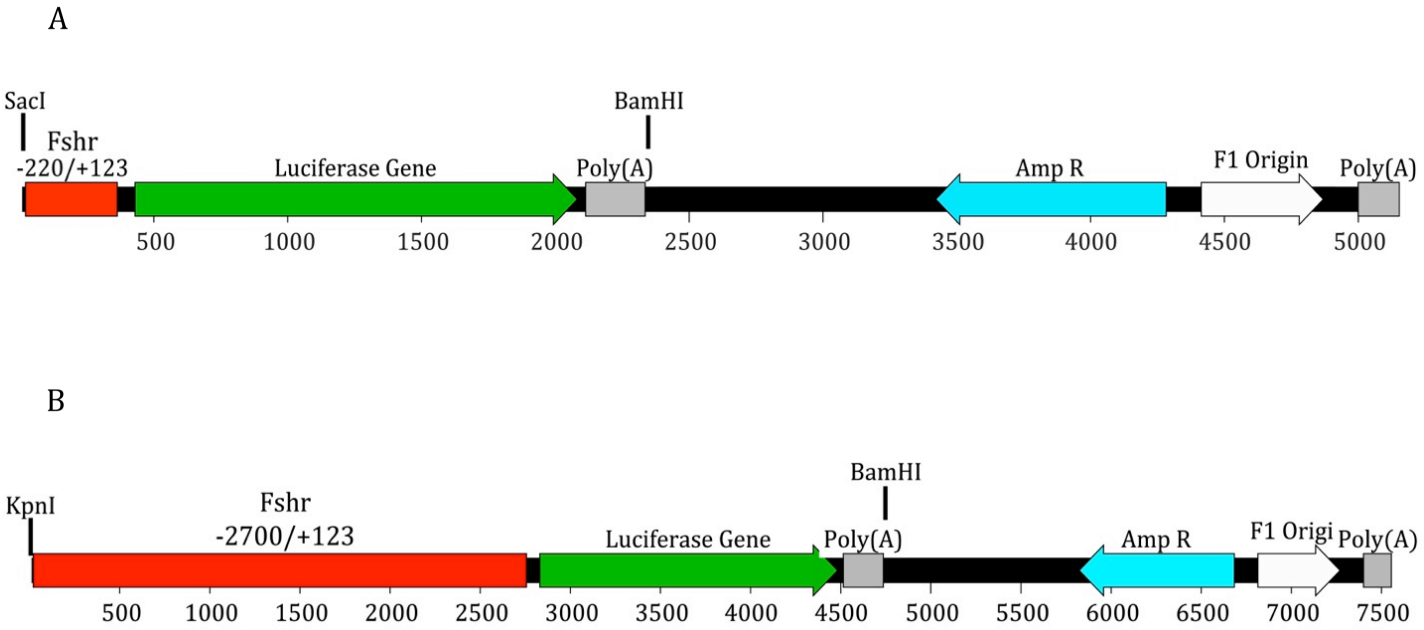



**Figure 4-1. Cloning vectors for transient transfections.** Vector construction for transient transfections containing (A) -200/+123 and (B) -2700/+123 region of the *Fshr* promoter in red, *Luciferase* gene in green and PolyA tails in grey. Vectors contained an *Amp R* gene for selection depicted in light blue and an F1 origin site for replication depicted in white. SacI, KpnI and BamHI insert sites are marked.

**Sertoli and Myoid cell Transfections.** Sertoli and myoid cells were transfected as previously described (Karl and Griswold 1990; Heckert, Daggett et al. 1998; Chen and Heckert 2001; Lei and Heckert 2002). Sertoli cells were plated directly into 24 well plates at a density of ~ 1.6 mg testes (based on the starting testes weight) per well. Forty-eight hours after plating, the cells were shocked with 10 mM Tris (pH 7.4) for two minutes to remove germ cells, and allowed to recover for at least 2 hours in fresh culture media. Immediately prior to transfection, the culture media was removed and 400 μl Dulbecco's Modified Eagle Medium (Sigma-Aldrich, St. Louis, MO), without added supplements, was added to each well. For each well, 300 ng plasmid DNA and 10 ng Renilla (pHRL-TK) was incubated with 1.5 μl Lipofectamine and 2ul Plus Reagent three hours at 37°C. Cells were fed 3 hours later with 300 μl straight DMEM. Prior to transfection, myoid cells were passaged twice and then plated in 24 well plates at a density of 52,000 cells per well. Eighteen hours after plating the cell media was changed to straight DMEM and cells were transfected using 100 ng plasmid DNA, 10 ng Renilla (pHRL-TK), 1ul Lipofectamine and 1ul Plus per well and incubated overnight 37°C, 18 hours later the transfection media was removed and the cells were fed with 1 ml complete media. Seventy-two hours after transfection, cells were lysed and assayed using the Dual-Luciferase® Reporter Assay System (Promega, Madison) for both firefly and *Renilla* following the manufactures instructions.

**Nuclear Extract Preparation.** Rat Sertoli cell nuclear extracts were prepared from rat Sertoli cells isolated as described in Chapter 2. Cells were washed with ice-cold

PBS, followed by a wash with five milliters of HEGD (25 mM HEPES pH 7.4, 1 mM DTT, 1.5mM EDTA pH 8.0, 10 % glycerol, 0.5 mM PMSF). Pellet resuspension was in in 10 mL HEDG for Sertoli cells and 6 mL HEDG for myoid cells. Cell lysis was achieved using an ice-cold Dounce homogenizer- 30 strokes with a type B pestle. Cellular suspensions were transferred to cold 1.5 mL microcentrifuge tubes and nuclei pelleted via centrifugation at 16,000 x g for one minute at 4°C. Supernatants were removed and nuclear pellets were resuspended in ice-cold HEGDK (25 mM HEPES pH 7.9, 1 mM DTT, 1.5mM EDTA pH 8.0, 10 % glycerol, 0.5M KCl, 0.5 mM PMSF) at a 1:1 ratio. Suspensions were mixed with a pipette tip, vortexed, and incubated on ice for one hour. Nuclear debris was removed by centrifugation at 100,000 x g for six minutes at 4°C. Supernatants were aliquoted and stored at -80°C. Protein concentration was determined by Bradford assays (BIO-RAD, Hercules, CA) following the manufacturers protocol.

**DNase I Footprinting Analysis.**

*Probe preparation*

DNase I footprinting was performed on a 5′-$^{32}$P-labeled PCR amplicon (Figure 4-2). Adenosine 5'-triphosphate [γ-$^{32}$P] label was added to PCR primers (Table 4-2) using T4 polynucleotide kinase (Thermo EK0031) as per manufacturers instructions. 40 pmol forward primer was incubated in 10 μl PNK buffer (500 mM Tris-HCl (pH 7.6 at 25°C), 100 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine) containing 50 μCi adenosine 5'-triphosphate [γ-$^{32}$P] and 10 U Polynucleotide kinase for 30 minutes at 37°C followed by a five minute incubation at 95°C. PCR amplification

with a 5' radiolabelled and an unlabeled 3' primer generated probes for Footprinting. PCR reactions (50μl) contained: 1x OptiBuffer (Bioline, Taunton, MA), four picomoles of radiolabelled forward primer, 12.5 pmol reverse primer, 50mM $MgCl_2$, 25mM dNTP's, 50 ng plasmid DNA and four units BIO-X-ACT Long DNA Polymerase (Bioline, Taunton, MA-#BIO-21049). Reactions were denatured for five minutes at 95°C followed by 35 amplification cycles of 95°C, 30 s; 56°C, 30 s; 72°C 30 s. A PCR cleanup kit (Qiagen, Germantown, MD -# 28104) purified radiolabelled PCR products. Specific activity was measured using a scintillation counter and probes were diluted to a final 'concentration' of ~50,000 cpm/μl.

*Ladder preparation*

The Sequenase Version 2.0 DNA Sequencing Kit (Affymetrix, Santa Clara, CA -# 70770) was used to prepare DNA sequencing ladders as per manufacturers instructions. 0.5 pmol plasmid DNA containing the rat ECR 1f sequence was denatured and incubated with 0.5 pmol forward primer in 10μl (fv) Sequenase reaction buffer (200mM Tris-HCl pH.7.5, 100mM $MgCl_2$, 250mM NaCl). Primer annealing was achieved by heating DNA/primer mix to 95°C for five minutes followed by five minutes at 56°C in a PCR machine. Annealed template was placed directly on ice and used immediately for ladder generation. 0.1 M DTT, labeling mix, 5 μCi [α-$^{32}$P]dATP and Sequenase Polymerase were incubated for two minutes at room temperature. 3.5 μl was transferred to new tubes containing termination mix (dideoxy nucleotide) and incubated for five minutes at 37°C. Four microliters of stop solution was added immediately to terminate the reaction and the samples were frozen at -20°C until needed. The Sequenase kit (Affymetrix, Santa Clara, CA)

contained all necessary buffers and enzymes. For electrophoresis, ladder preps were heated to 75°C for five minutes prior to loading.

*DNaseI Digestion - Footprinting*

Eighty microliters of reaction buffer (10mM HEPES pH 7.9, 40 mM KCl, 2 mM $MgCl_2$, 5% glycerol) containing 20mM DTT, 1 μg poly(dI:dC), labeled probe (~50,000 CPM), nuclear extract and DNaseI (Worthington, Lakewood, NJ - #LS006333) was used for each footprinting reaction. DNase concentration, nuclear extract concentration and digest time varied, see Table 4-3. Samples were incubated at room temperature for the allotted time and digests were stopped by the addition of 200 μl Tris-saturated Phenol and vortexed on max setting for 10 seconds. 120μl water was added, vortexed and samples were centrifuged for 10 minutes at 16,000 x g. The aqueous phase was transferred to a clean tube and re-extracted with a 1:1 mixture of phenol:chloroform, twice. A final chloroform extraction of the aqueous phase was preformed. The final DNA supernatant was removed and transferred to a clean tube. 300mM Sodium acetate (pH 5.2) was added and the samples were vortexed. One milliliter of ice-cold 100% Ethanol was added to each sample, thoroughly mixed and incubated at -80°C for 15 minutes. DNA was pelleted by centrifugation at 16,000 x g for 20 minutes at 4°C, washed with 60% Ethanol, air-dried and resuspended in 5 μl 1x stop solution (Sequenase Kit). Samples were heated to 95°C for 5 minutes prior to loading on the sequencing gel.

*Sequencing Gel and Electrophoresis*

The DNA samples were resolved by denaturing PAGE and visualized by autoradiography. A denaturing 8% polyacrylamide (7M Urea) gel was cast and

allowed to polymerize overnight at room temperature. The gel was pre-ran in 1x TBE until the gel temperature reached 45°C. Resolution of the digested DNA samples was achieved by electrophoresis for 2.5 hours at 70 watts. Gel temperature remained around 45°C the entire run. The gel was transferred to filter paper and vacuum dried for two hours at 50°C. The dried gel was exposed to autoradiography film at room temperature.

**Table 4-2. PCR primers used for footprinting probe generation**.

| Primer | Sequence |
|---|---|
| 1fa1 F | GGCCATGCGGAGCTCGAGCCTGTTTCTTGGATGCC |
| 1fa1 R | CTCATCACAGGAATGCCTTCTG |
| 1fa2 F | GATACTCCATTCTTCCTGCAAAGG |
| 1fa2 R | AGGCGTGAGGATAGGCTTTTAA |
| 1b F | TCTTGGGGCTCAGTGTGGAGT |
| 1b R | GTCGTGGATCCGAAGTAAAAATACTC |

**Table 4-3. DNaseI concentration and digestion times for footprinting reactions**.

| Nuclear Extract | Concentration (μg) | DNaseI Concentration (U) | Digest Time (m:s) |
|---|---|---|---|
| Sertoli | 30 | 0.2 | 2:30 |
| Myoid | 30 | 0.1 | 1:15 |
| TM4 | 30 | 0.1 | 1:10 |
| MSC1 | 30 | 0.1 | 1:10 |
| **Probe only** | | | |
| 1f1 | - | 0.05 | 22s |
| 1f2 | - | 0.05 | 22s |
| 1b | - | 0.05 | 22s |

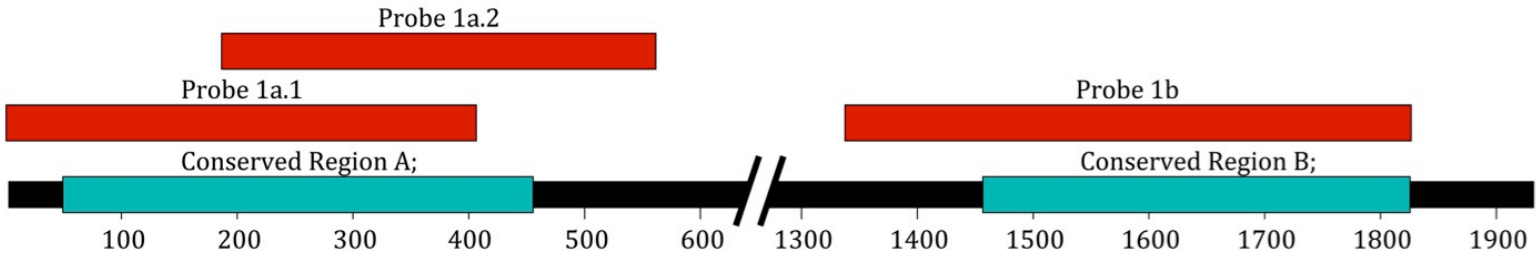



**Figure 4-2. Annotated depiction of probes used for DNaseI footprint analysis.** The Red horizontal bars indicate genomic location of footprint probes generated by PCR. Conserved regions ECR 1f-a and 1f-b are indicated by the blue horizontal bars. The ruler under the figure represents genomic size in base pairs. Sequence between 650 bp and 1250 bp was removed to reduce image size.

**Electric Mobility Shift Assays.** EMSA assay were preformed as previously described with slight modifications (Daggett, Rice et al. 2000; Lei and Heckert 2002). The LUEGO (Labeled Universal Electrophoretic Gel shift Oligonucleotide method was used to add a 5' fluorescent label on all EMSA probes (Jullien and Herman 2011). Nuclear extract, competitors and antibodies were incubated for 20 minutes on ice in EMSA buffer (20 mM HEPES pH 7.9, 80 mM KCl, 4 mM $MgCl_2$, 10% glycerol, 250 μM $ZnCl_2$, 0.5 mM DTT, 200 ng ssDNA, 500 ng poly(dI-dC)). Cold competitors (100 fold molar excess) and antibody (2 μg) were included prior to addition of nuclear extracts (Table 4-4). The following antibodies were purchased from Santa Cruz Biotechnology, Inc (Santa Cruz, CA): Goat polyclonal anti-GATA-1(M-20) IgG, mouse monoclonal anti-GATA-2 (CG2-96) IgG, mouse monoclonal anti-GATA-3 (HG3-31) IgG, rabbit polyclonal anti-GATA-4 (H-112) IgG, rabbit polyclonal anti-GATA-6 (H-92) IgG.

**Table 4-4. Oligonucleotides used for EMSA probes and competitors.**

| Probe | Sequence 5' to 3' |
|---|---|
| LUEGO 5' IR-800 | /5IRD800/GTGCCCTGGTCTGG |
| ECR 1f 3.2-1 S | TGATAGGAGAAACCACAATCCAAAACTATGTTCCT |
| ECR 1f 3.2-1 AS | AGGAACATAGTTTTGGATTGTGGTTTCTCCTATCACCAAGACCAGGGCAC |
| ECR 1f 3.2-2 S | AACTATGTTCCTTTTTCAGAAGCAAAACCACAGAA |
| ECR 1f 3.2-2 AS | TTCTGTGGTTTTGCTTCTGAAAAAGGAACATAGTTCCAAGACCAGGGCAC |
| ECR 1f 3.3-1 S | GCAAAACCACAGAAGGCATTCCTGTGATGAGTTCA |
| ECR 1f 3.3-1 AS | TGAACTCATCACAGGAATGCCTTCTGTGGTTTTGCCCAAGACCAGGGCAC |
| ECR 1f 3.3-2 S | CCTGTGATGAGTTCAGACATTATTCATAAGGCGTT |
| ECR 1f 3.3-2 AS | AACGCCTTATGAATAATGTCTGAACTCATCACAGGCCAAGACCAGGGCAC |
| **Competitor** | **Sequence 5' to 3'** |
| ECR 1f 3.2-1 mut1 S | TGA**G**AGGAGAAACCACAATCCAAAACTATGTTCCT |
| ECR 1f 3.2-1 mut1 AS | AGGAACATAGTTTTGGATTGTGGTTTCTCCT**C**TCA |
| ECR 1f 3.2-1 mut2 S | **ACGCCT**GAGAAACCACAATCCAAAACTATGTTCCT |
| ECR 1f 3.2-1 mut2 AS | AGGAACATAGTTTTGGATTGTGGTTTCTC**AGGCGT** |
| ECR 1f 3.2-1 mut3 S | TGATAGGAGAAAC**GTAGCAACG**AAACTATGTTCCT |
| ECR 1f 3.2-1 mut3 AS | AGGAACATAGTTT**CGTTGCTAC**GTTTCTCCTATCA |
| ECR 1f 3.2-1 mut4 S | TGATAGGAGAAACCACAATCCAAAAC**CGGA**TTCCT |
| ECR 1f 3.2-1 mut4 AS | AGGAA**TCCG**GTTTTGGATTGTGGTTTCTCCTATCA |
| ECR 1f 3.3-2 mut 1 S | **AGACCT**ATGAGTTCAGACATTATTCATAAGGCGTT |
| ECR 1f 3.3-2 mut 1 AS | AACGCCTTATGAATAATGTCTGAACTCAT**AGGTCT** |
| ECR 1f 3.3-2 mut 2 S | CCTGTG**GCAGTG**TCAGACATTATTCATAAGGCGTT |
| ECR 1f 3.3-2 mut 2 AS | AACGCCTTATGAATAATGTCTGA**CACTGC**CACAGG |
| ECR 1f 3.3-2 mut 3 S | CCTGTGATGAGT**GGCTAG**ATTATTCATAAGGCGTT |
| ECR 1f 3.3-2 mut 3 AS | AACGCCTTATGAATAAT**CTAGCC**ACTCATCACAGG |
| ECR 1f 3.3-2 mut 4 S | CCTGTGATGAGTTCAGAC**CGTATC**CATAAGGCGTT |
| ECR 1f 3.3-2 mut 4 AS | AACGCCTTATG**GATACG**GTCTGAACTCATCACAGG |
| ECR 1f 3.3-2 mut 5 S | CCTGTGATGAGTTCAGACATTATT**TGAACT**GCGTT |
| ECR 1f 3.3-2 mut 5 AS | AACGC**AGTTCA**AATAATGTCTGAACTCATCACAGG |
| ECR 1f 3.3-2 mut 6 S | CCTGTGATGAGTTCAGACATTATTCATAAG**CTAGG** |
| ECR 1f 3.3-2 mut 6 AS | **CCTAG**CTTATGAATAATGTCTGAACTCATCACAGG |
| GATA consensus oligo S | CACTTGATAACAGAAAGTGATAACTCT |
| GATA consensus oligo AS | AGAGTTATCACTTTCTGTTATCAAGTG |

Bold font indicates mutated sequence and underlined text indicates consensus sequence

# Results

**Conservation profile of the *Fshr* locus.** To identify candidate distal regulatory elements, a 1.4 Mb region of genomic DNA that included *Fshr* and its intergenic sequences was analyzed for sequence conservation and 7x regulatory potential, using the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway) and associated software(Kolbe, Taylor et al. 2004; King, Taylor et al. 2005; Taylor, Tyekucheva et al. 2006). Evolutionarily conserved sequences were identified by comparing the sequence of several species of divergent evolutionary origins, using *Homo sapiens* as the base genome and *Gallus gallus* (chicken) as the most distant. To identify DNA sequences with a high probability of regulatory potential, only conserved regions with an increased 7x regulatory score were selected. The 7X regulatory potential scores were computed from human, chimpanzee, macaque, mouse, rat, dog, and cow by comparing short alignment pattern frequencies between known regulatory elements and neutral DNA (Waterston, Lindblad-Toh et al. 2002; Consortium 2004; Gibbs, Weinstock et al. 2004; Havlak, Chen et al. 2004; Kolbe, Taylor et al. 2004; King, Taylor et al. 2007). This allows selection of regions that not only are highly conserved, but also contain DNA sequences that have demonstrated regulatory activity associated with other genes.

The analysis identified 30 evolutionarily conserved regions (ECR's) that also showed sequence similarity to known regulatory elements, which were further considered as possible *Fshr* regulatory regions. The ECR's were then annotated on the *Rattus norvegicus* genome listed in Table 4-5 and shown graphically in Figure 4-

3. Seventeen ECR's mapped to the 5' region of *Fshr*, seven to the intergenic region, and six were located 3' of the gene. The rat genome was selected for further study, as Sertoli cell isolation, transcriptional assays and promoter analysis methods are firmly established for this species. Thirty ECRs were then evaluated to determine their role in *Fshr* regulation.

**Table 4-5. Genomic location of ECR's in the *Rattus norvegicus* genome.**

| ECR | Chromosome Location | Size (bp) |
|---|---|---|
| ECR 1 | Chr6: 23036890-23037477 | 588 |
| ECR 1c | Chr6: 22585376-22586088 | 713 |
| ECR 1d | Chr6: 22597052-22597262 | 211 |
| ECR 1f | Chr6: 22614761-22616669 | 1909 |
| ECR 1g | Chr6: 22590321-22590502 | 182 |
| ECR 2 | Chr6: 22626739-22629962 | 3224 |
| ECR 3 | Chr6: 22596865-22597324 | 460 |
| ECR 4 | Chr6: 23287153-23287391 | 239 |
| ECR 5 | Chr6: 23423202-23423600 | 399 |
| ECR 6 | Chr6: 23219667-23220065 | 399 |
| ECR 7 | Chr6: 23437816-23438244 | 429 |
| ECR 8 | Chr6: 22771058-22771652 | 595 |
| ECR 9 | Chr6: 22776439-22777106 | 668 |
| ECR 9b | Chr6: 22785969-22786300 | 332 |
| ECR 10 | Chr6: 22813283-22813805 | 523 |
| ECR 11 | Chr6: 22826398-22826718 | 321 |
| ECR 12 | Chr6: 22826285-22826757 | 473 |
| ECR 13 | Chr6: 22845802-22846436 | 635 |
| ECR 14 | Chr6: 23024900-23025427 | 528 |
| ECR 15 | Chr6: 23048424-23049218 | 795 |
| ECR 16 | Chr6: 23055601-23056494 | 894 |
| ECR 17 | Chr6: 23148636-23149418 | 783 |
| ECR 18 | Chr6: 23153766-23154371 | 606 |
| ECR 19 | Chr6: 23167660-23168222 | 563 |
| ECR 20 | Chr6: 23173974-23174670 | 697 |
| ECR 21 | Chr6: 23180610-23181272 | 663 |
| ECR 22 | Chr6: 23188053-23188745 | 693 |
| ECR 23 | Chr6: 23195435-23196040 | 606 |
| ECR 24 | Chr6: 23212514-23213040 | 527 |
| ECR 25 | Chr6: 23259302-23260347 | 1046 |

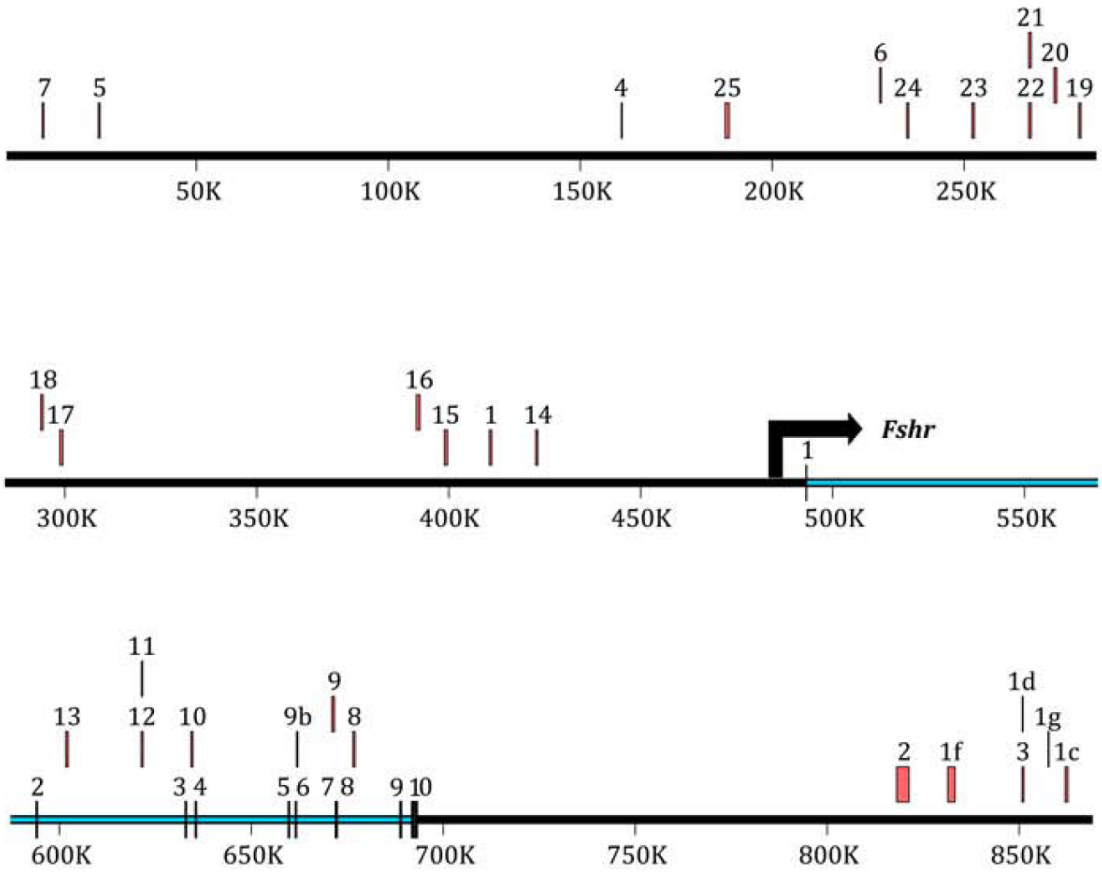



**Figure 4-3. ECRs identified by sequence conservation and regulatory potential annotated on the *Rattus norvegicus* genome**. Sequence identified by comparative genomics were identified within the human genome and transposed to the rat genome using the liftover tool in the UCSC genome browser. The light blue line denotes the *Fshr* gene and the black arrow indicates the direction of transcription. Solid horizontal black lines within the gene identify *Fshr* introns and are numbered in black just above. ECR's are identified by the solid horizontal red bars above the sequence and are labeled accordingly.

**H3K4me3 is a highly enriched histone modification of ECR 1f**. To prioritize our selection of ECR's, Chromatin Immunoprecipitation (ChIP) was coupled with deep sequencing to help identify regions enriched for Histone 3 lysine 4 tri-methylation (H3K4me3) modification in Sertoli and myoid cells. H3K4me3 enrichment is commonly associated with active promoter and enhancer elements; as such we used Sertoli cells to identify active regions within the *Fshr* locus and myoid cells as a representative inactive locus. Sequenced ChIP DNA was aligned to version 5.0 of the *Rattus norvegicus* genome assembly. H3K4me3 Enrichment is indicated by the peak height, which is a direct reflection of the number of sequence tags identified. Peak finding analysis also identified ECR 1f as the most enriched site in the Sertoli cell genome (Chapter 3), which can be appreciated when viewing the enrichment profiles in the UCSC Genome Browser across the *Fshr* locus (Figure 4-4).

Overall, ECR 1f was the only ECR associated with a histone profile indicative of activity in Sertoli cells. Enrichment of H3K4me3 is higher in the Sertoli cell fraction within ECR 1f across the entire ECR. Notably, the enrichment was not observed in myoid cells, suggesting the region's cell-specific activity within Sertoli cells, the active cell type.

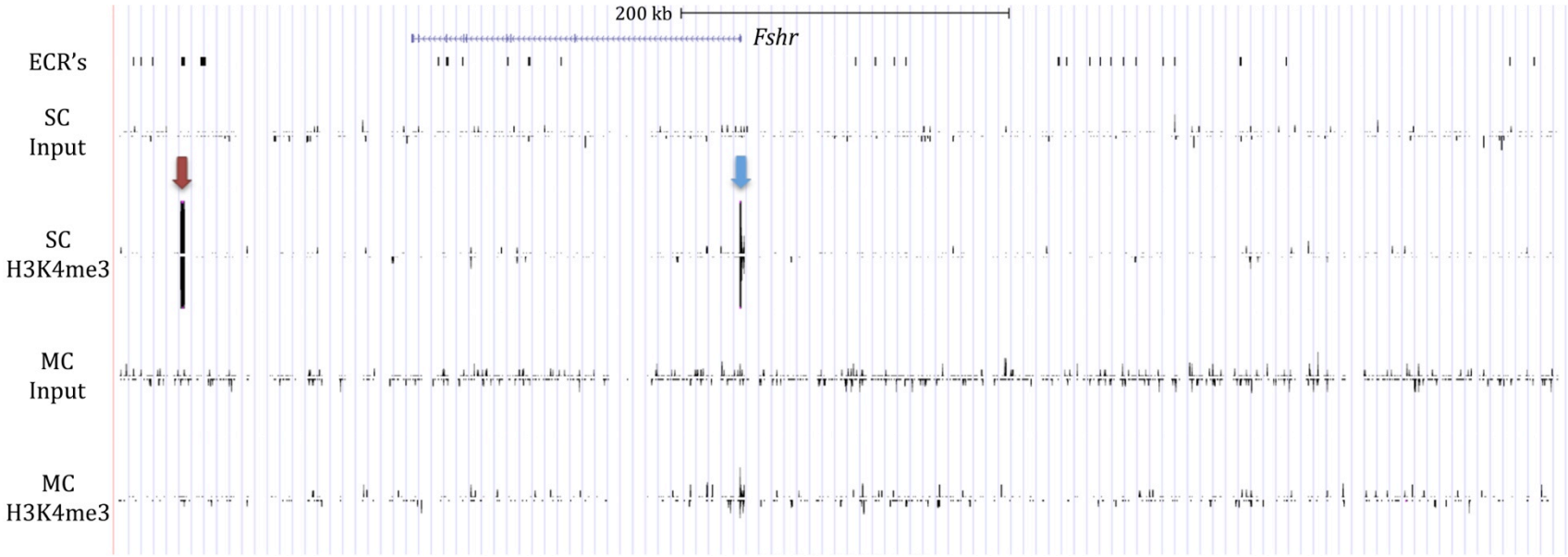



**Figure 4-4. Enrichment profile of H3K4me3 across the *Fshr* locus.** Histograms generated on the UCSC genome browser display the enrichment profiles of Input and H3K4me3 samples from both Sertoli and myoid cells. Scale bar= 200 kb, *Fshr* is denoted by the purple bar and ECR's are identified by black vertical bars. Enrichment peaks were identified for the *Fshr* promoter (blue arrow) and the ECR 1f (red arrow).

**ECR 1f is associated with other histone modifications indicative of active chromatin.** To further understand the chromatin state within ECR 1f, ChIP coupled with qPCR was employed to determine other histone modifications and DNA binding proteins present in freshly isolated P15 Sertoli cells *in vivo*. As seen in the ChIP-Seq data, there was an abundance of H3K4me3 associated with ECR 1f (Figure 4-5). In addition, histone 4 acetylation (H4ac), a mark of active chromatin, was enriched in ECR 1f, while Histone 3 lysine 27 tri-methylation (H3K72me3) and Histone 3 lysine 9 tri-methylation (H3K9me3), markers of repressed chromatin, were not. This data confirms our ChIP-Seq results and shows that ECR 1f is associated with other active, but not repressive, chromatin modifications.

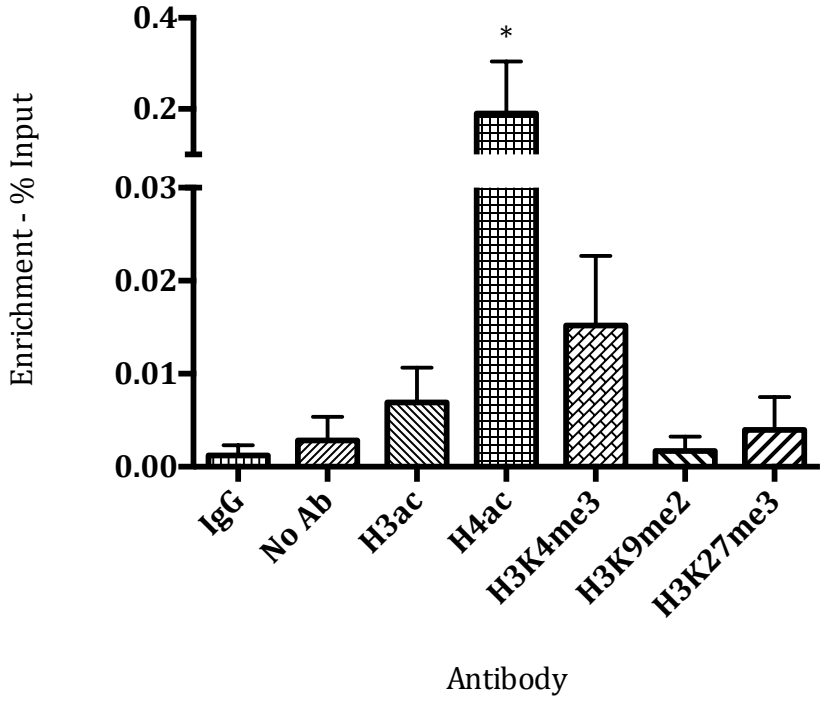



**Figure 4-5. ChIP-qPCR verification of ChIP-Seq and identification of other associated histone modifications.** ChIP in P15 Sertoli cells identifying protein enrichment at ECR 1f. Histone modifications selected for are: H3ac, H4ac, H3K4me3, H3K9me2, H3K27me3, IgG and a No Antibody (No Ab) control. Enrichment values are expressed as a percent of the Input sample. Error bars reflect the standard deviation. Statistically significant differences are indicated by * (p-value≤0.05) as determined by one-way ANOVA.

**ECR 1f regulates *Fshr* promoter activity.** To determine if ECR 1f regulates *Fshr* promoter activity, it was cloned into a firefly luciferase reporter together with the -220/+123 and -2700/+123 *Fshr* promoter sequences, and transiently transfected into primary rat Sertoli and myoid cells. ECR 1f was inserted into the reporter vector both 5' and 3' to the promoters and in both forward and reverse orientations. Firefly luciferase activity in the transfected cells was normalized to the activity of *Renilla* luciferase generated by the co-transfected tK-REN vector included as a control for transfection efficiency. The Firefly/*Renilla* values from cells transfected with ECR 1f reporters were then normalized to the values from cells transfected with their respective promoter-only vectors. This revealed an approximate four-fold reduction in promoter activity in the presence of ECR 1f in both promoter constructs tested (Figure 4-6). ECR 1f activity was independent of orientation (forward or reverse), position (5' or 3' relative to *luciferase*), promoter size (-220/+123 or -2700/+123) or cell type. These results indicate that ECR 1f contains sequences that can regulate *Fshr* promoter activity in transfections in both Sertoli and myoid cells and this activity is independent of promoter size and element position. This observation conflicts with our ChIP-Seq data that identified ECR 1f as a cell-specific regulatory region. This could imply that the DNA binding proteins responsible for ECR 1f's activity are not cell specific; as this is a transient assay no chromatin modifications exist to limit the activity of ECR 1f within our reporter in myoid cells. All other ECRs identified by comparative genomics were also tested by transient transfection analysis and were unable to alter *Fshr* promoter activity *in vitro* (data not shown).

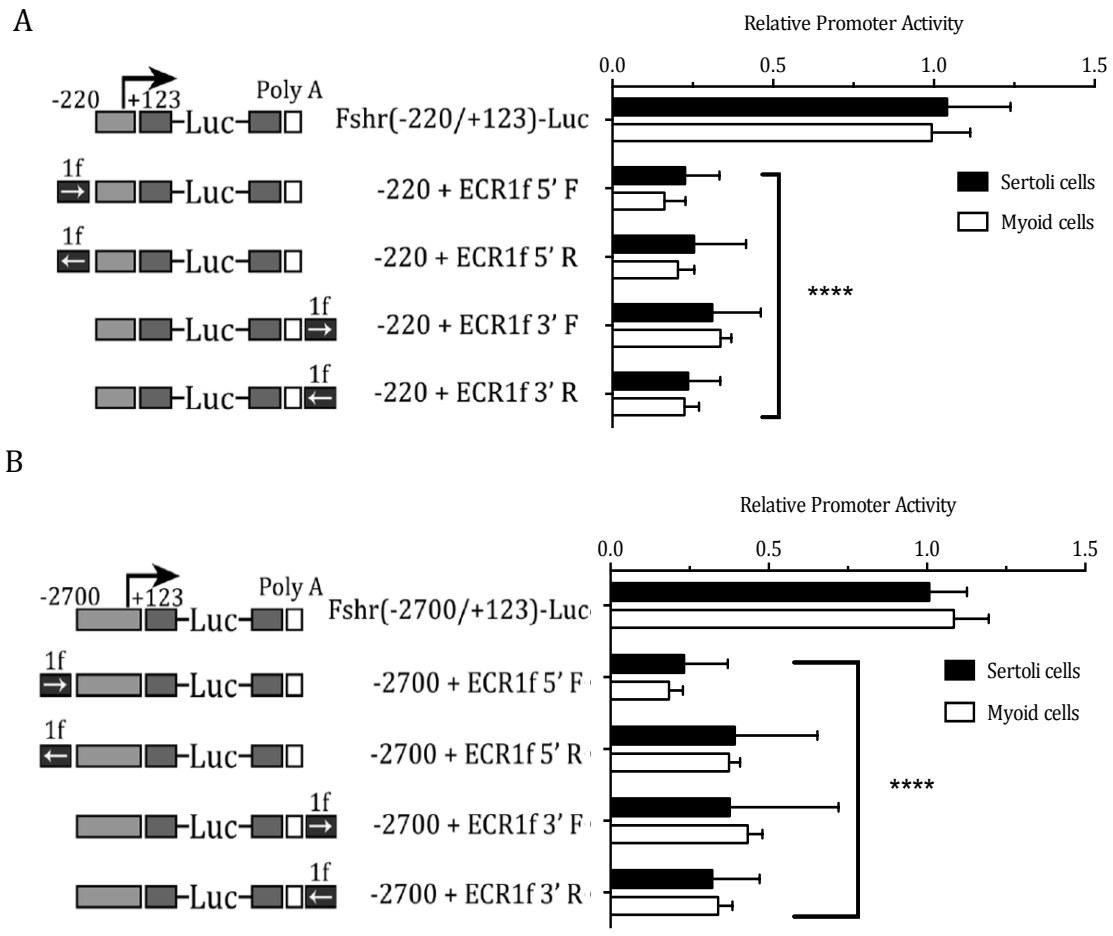



**Figure 4-6. ECR 1f represses *Fshr* promoter activity in transient transfections.** Plasmids containing ECR 1f were inserted into the pGem vector containing the (A) -200/+123 bp *Fshr* promoter and (B) -2700 /+123 bp *Fshr* promoter were transiently transfected into primary Sertoli (black bars) cells and myoid (white bars) cells. A schematic of the transfection vector represents the plasmid construct used for each transfection. The light grey box represents the *Fshr* promoter and the arrow indicates transcriptional direction, the dark grey box represents the *Luciferase* gene, the white box represents the polyA tail and the black box with an arrow represents the inserted ECR and the direction of the inserted sequence. Error bars reflect the standard deviation. Statistically significant differences are indicated by **** (p-value≤0.0001) as determined by two-way ANOVA.

**Identification of two highly conserved regions within ECR 1f.** To better define the active sequences within the 2000 bases of ECR 1f, conservation was used to guide the generation of reporter with small portions of ECR 1f directing *Fshr*. Detailed inspection of the sequence conservation in ECR 1f revealed two distinct regions ECR 1f-a and ECR 1f-b (Figure 4-7). The conservation within these two regions exists in the chicken, but no conservation exists with Zebrafish or *Xenopus* (data not shown). ECR 1f-a is located at the 5' region of ECR 1f and is roughly 400 base pairs, while ECR 1f-b is in the 3' region of ECR 1f and is roughly 370 base pairs. The two conserved regions are separated by about 1000 bp of less-conserved sequence.

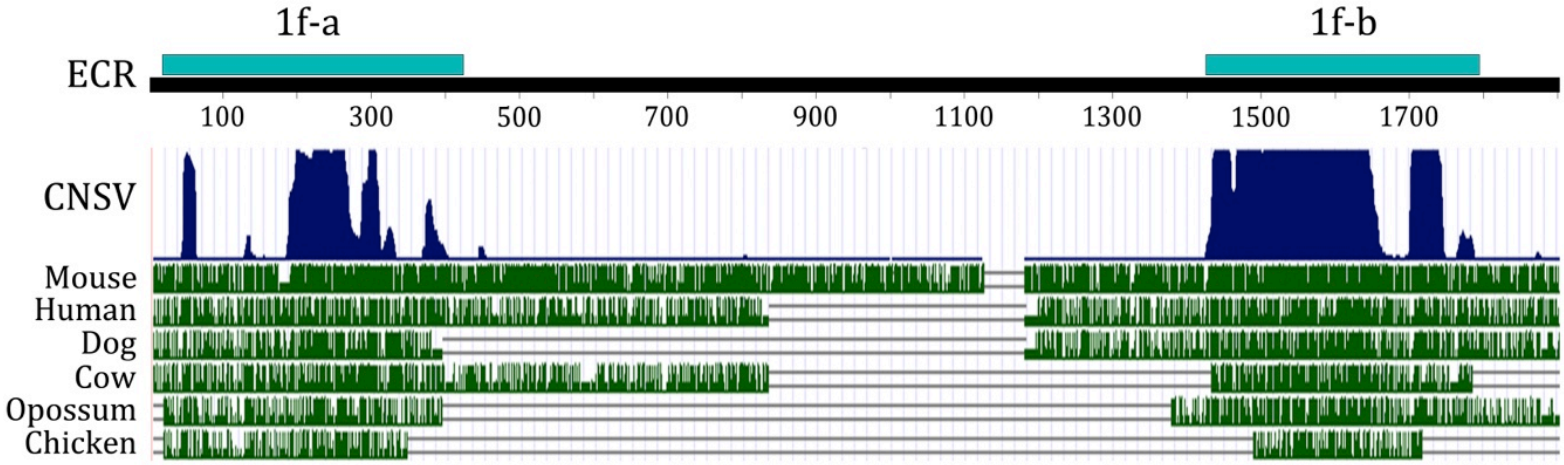



**Figure 4-7. Sequence conservation within ECR 1f.** Species-specific sequence conservation (green histograms) and vertebrate consensus sequence conservation (blue histogram-CNSV) were visualized across ECR 1f (black horizontal bar) using the UCSC Genome Browser. This revealed two distinct regions of conservation ECR 1f-a and ECR 1f-b (horizontal teal bars).

**Both ECR 1f-a and 1f-b are required for full ECR 1f activity.** To determine the effect of ECR 1f-a and 1f-b on *Fshr* promoter activity, the two were individually cloned into the FSHR -220/+123 and FSHR -2700/+123 vectors and transiently transfected into P15 rat Sertoli and myoid cells (Figure 4-8). Two significant trends were associated their promoter activity. First, ECR 1f-a enhanced transcription in myoid cells, in a manner that depended on its orientation (forward) and position (3'). Second, ECR 1f-b repressed transcription in both Sertoli and myoid cells; but only when it was at the 3' end in reverse orientation. Similar results were seen with both the -220 and -2700 promoters. Since the enhancer activity is only found in myoid cells, it is likely that the protein involved in this activation is only present in myoid cells and is not likely to play a role in *Fshr* repression *in vivo.* The repressive action associated with ECR 1f-b indicates function in both Sertoli and myoid cells, but is only associated with activity in the reverse orientation when located 3' of the reporter gene. This position is consistent with ECR 1f *in vivo* as ECR 1f is downstream of *Fshr.* Neither ECR 1f-a or 1f-b alone recapitulates the activity of the full ECR 1f.

As distal regulatory elements often work in concert with each other, ECR 1f-a and 1f-b were joined and cloned into the FSHR -220/+123 vector in the 5' forward direction and transiently transfected into P15 rat Sertoli and myoid cells. Together these two segments are capable of repressing the *Fshr* promoter (Figure 4-9), although not to the same level as the full length ECR. This observation indicates that required functional elements are contained within each conserved region of ECR 1f,

but that the non-conserved, intervening sequence also plays a role in the function of ECR 1f.

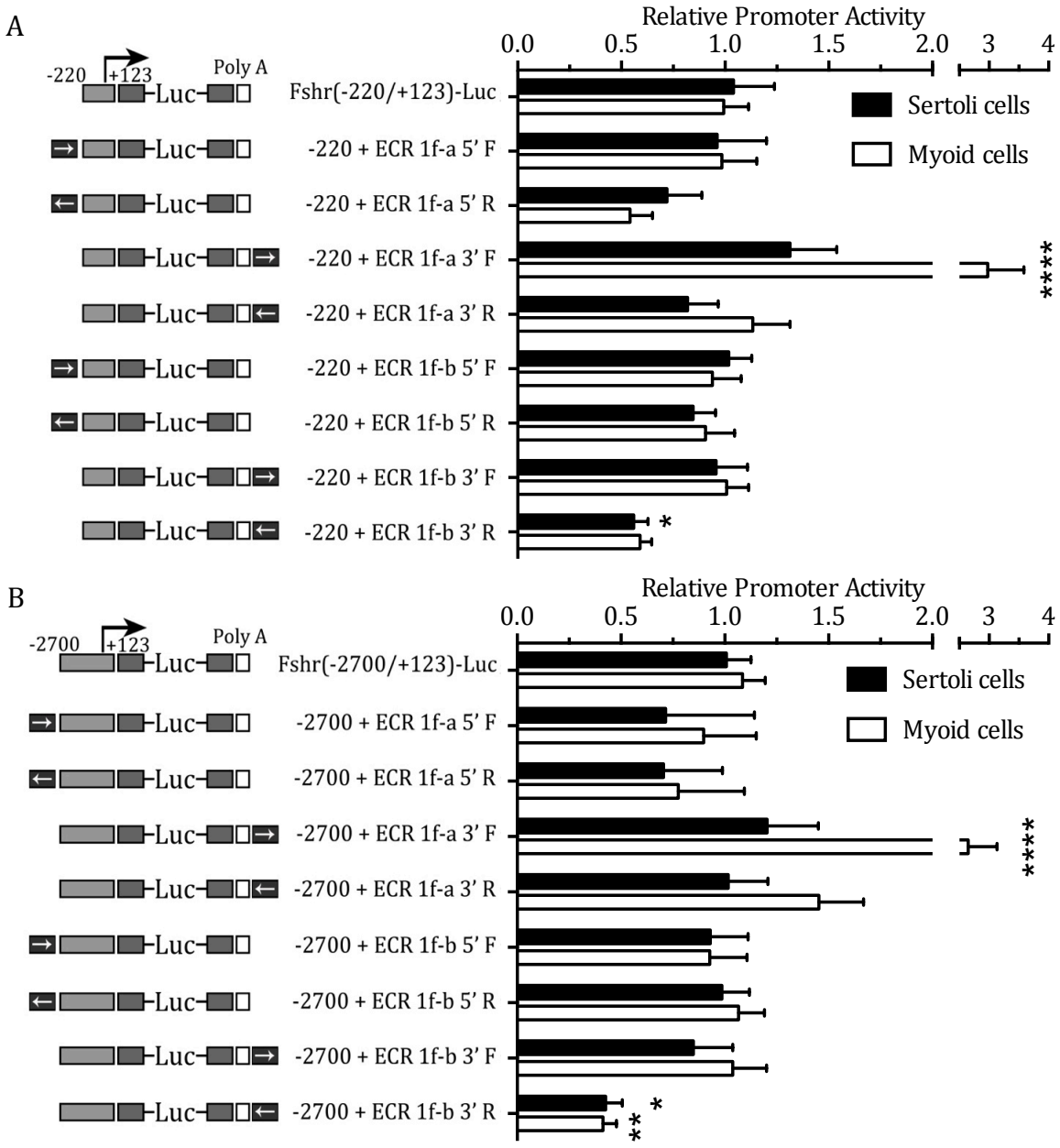



**Figure 4-8. Transient transfection analysis of ECR 1f-a and 1f-b in Sertoli and myoid cells.** Plasmids containing ECR 1f-a and ECR 1f-b were cloned into the pGem vector containing the (A) -200/+123 bp *Fshr* promoter and (B) -2700 /+123 bp *Fshr* promoter. A schematic of the transfection vector represents the plasmid construct used for each transfection. The light grey box represents the *Fshr* promoter and the arrow indicates transcriptional direction, the dark grey box represents the *Luciferase* gene, the white box represents the polyA tail and the black box with an arrow represents the inserted ECR and the direction of the inserted sequence. Error bars reflect the standard deviation. Statistically significant differences are indicated by * p-value≤0.05, ** p-value≤0.01, **** p-value≤0.0001 as determined by two-way ANOVA.

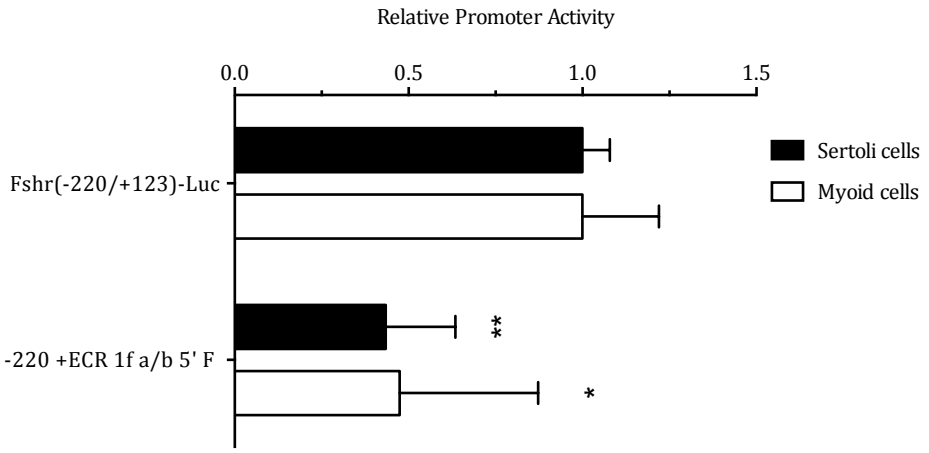



**Figure 4-9. Transient transfection analysis of ECR 1f-a/b in Sertoli and myoid cells.** Plasmids containing a fusion of ECR 1f-a/b was cloned into the pGem vector 5' of the -220/+123 bp *Fshr* promoter in the forward orientation. Black bars represent Sertoli cells and the white bar represents the myoid cells. Error bars reflect the standard deviation. Statistically significant differences are indicated by * p-value≤0.05, ** p-value≤0.01, as determined by two-way ANOVA.

***In Vitro* DNaseI Footprinting identified protein-binding sites within ECR1f-a and 1f-b.** To help reveal the active sequences in ECR 1f, *in vitro* DNase I Footprint analysis was used to identify sites of protein binding and/or hypersensitive regions within the conserved regions of ECR 1f. Three different probes (Figure 4-2) were used to cover the sequences within ECR 1f-a (1a.1 and 1a.2) and ECR 1f-b (1b), together with nuclear extracts from primary Sertoli and myoid cells as well as two Sertoli cell lines-TM4 and MSC1. The analysis identified 46 sequences that were protected by proteins present in the nuclear extracts (Table 4-6 and Figure 4-10) and their binding regions annotated across ECR 1f-a and ECR 1f-b (Figure 4-11 A and B, respectively). Only two footprints, footprint 13 and 18b, were specific to Sertoli cells and MSC1 cells, while nine footprints/hypersensitive sites were observed in both Sertoli and myoid cells that exhibited different banding patterns between Sertoli and myoid cells (asterisks, Table 4-5 and Figures 4-10 and 4-11).

Overall, more myoid cell-specific footprints were observed (13) and fifteen footprints were similar between both cell types. More footprints were observed in ECR 1f-a when compared to ECR 1f-b; likely because the probe for ECR 1f-b was too large and did not resolve well towards the 3' end of the probe. Although the intergenic sequence was not analyzed, our transient transfection data indicates that this region contains sequence required for full ECR 1f activity. This region should be analyzed in future experiments.

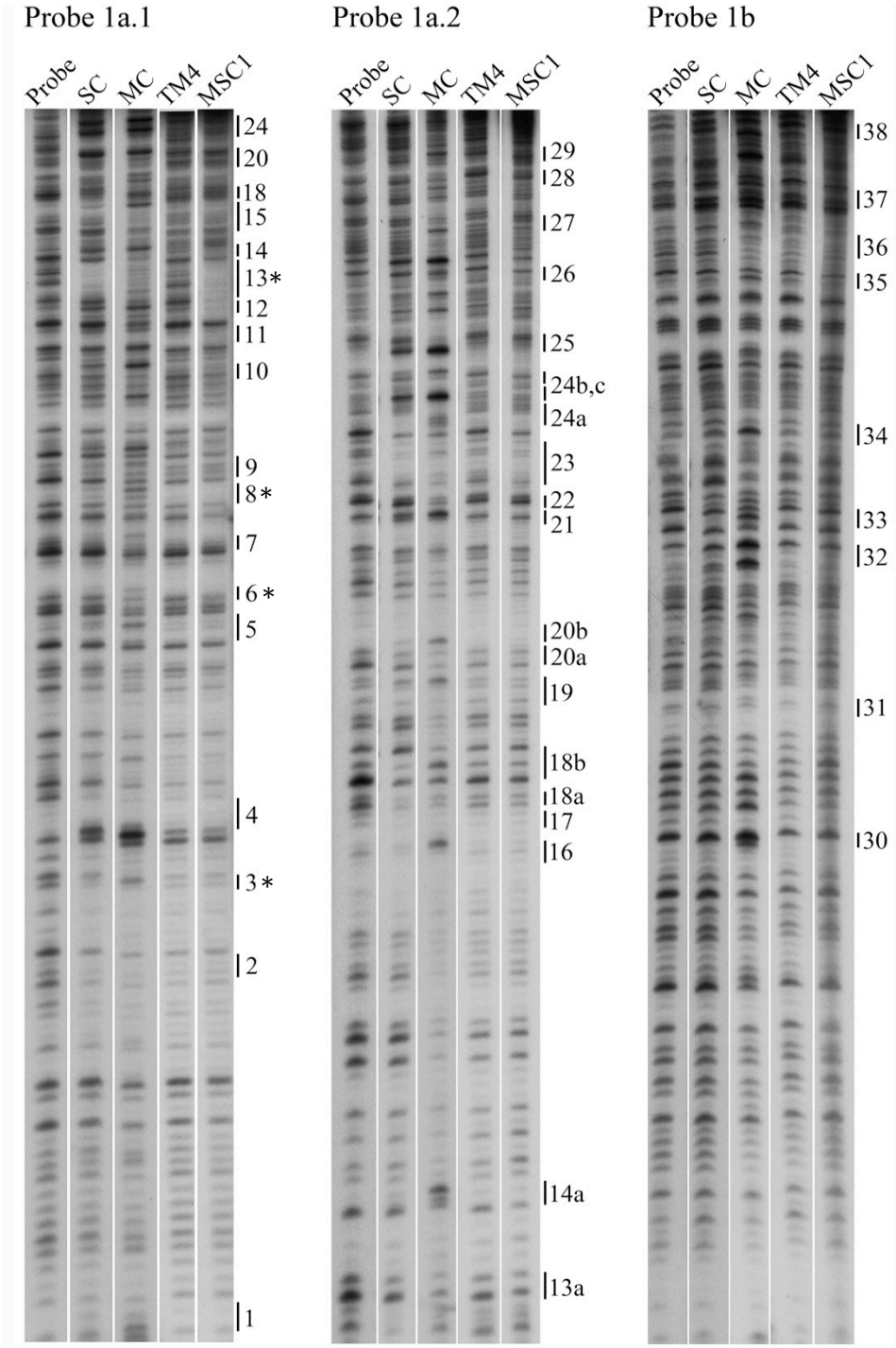



**Figure 4-10. 46 footprinted and hypersensitive regions within ECR 1f-a and 1f-b.** DNase I footprinting analysis of ECR1f-a and 1f-b with Probe only, Sertoli (SC), myoid (MC), TM4 and MSC1 nuclear extract. ECR 1f-a is covered by two probes- 1a.1 covers the 5' region and probe 1a.2 covers the 3' region. A single probe covers ECR 1f-b. * Indicates different footprint pattern between cell types.

**Table 4-6. Chromosome positions of footprint and hypersensitive sites.**

| FP Number | Probe | Chromosome Location | DNA Sequence | Cell Type |
|---|---|---|---|---|
| 1 | 1a1 | chr6:22614772-22614775 | acca | SC, MC |
| 2 | 1a1 | chr6:22614800-22614804 | tctct | All |
| 3* | 1a1 | chr6:22614808-22614812 | aagca | SC, MC |
| 4 | 1a1 | chr6:22614816-22614820 | gggag | All |
| 5 | 1a1 | chr6:22614845-22614849 | ggctc | MC |
| 6* | 1a1 | chr6:22614854-22614857 | caaa | MC |
| 7 | 1a1 | chr6:22614862-22614866 | ctgtg | MC |
| 8* | 1a1 | chr6:22614874-22614879 | tgaata | MC |
| 9 | 1a1 | chr6:22614881-22614885 | gcaac | MC |
| 10 | 1a1 | chr6:22614908-22614916 | ccctccca | SC, MC |
| 11 | 1a1 | chr6:22614925-22614931 | ttcttcc | MC |
| 12 | 1a1 | chr6:22614935-22614942 | aaaggct | MC, MSC1 |
| 13* | 1a1 | chr6:22614947-22614964 | ggaggctagataatatc | SC, MC, MSC1 |
| 13a | 1a2 | chr6:22614959-22614965 | agtatct | All |
| 14 | 1a1 | chr6:22614968-22614974 | gaggaaa | All |
| 14a | 1a2 | chr6:22614968-22614971 | gagg | SC, MC |
| 15 | 1a1 | chr6:22614981-22614992 | cctaattagcaa | SC, MC |
| 16 | 1a2 | chr6:22614997-22615000 | atta | SC, MC |
| 17 | 1a2 | chr6:22615002-22615005 | atta | SC, MC |
| 18 | 1a1 | chr6:22615006-22615015 | cacacatcct | SC, MC |
| 18a | 1a2 | chr6:22615005-22615009 | acaca | SC, MC |
| 18b* | 1a2 | chr6:22615008-22615011 | caca | SC, MC |
| 19 | 1a2 | chr6:22615016-22615020 | gacaa | SC, MC |
| 20 | 1a1 | chr6:22615024-22615034 | ccatttttaat | TM4, MSC1 |
| 20a* | 1a2 | chr6:22615023-22615025 | tcc | SC, MC |
| 20b* | 1a2 | chr6:22615027-22615031 | ttttt | SC, MC |
| 21* | 1a2 | chr6:22615045-22615047 | act | SC, MC |
| 22 | 1a2 | chr6:22615049-22615052 | gttg | MC |
| 23* | 1a2 | chr6:22615054-22615064 | ttaaatctga | SC, MC |
| 24 | 1a1 | chr6:22615066-22615092 | ggagaaaccacaaatcca aaactatgtt | SC, MC |
| 24a | 1a2 | chr6:22615068-22615073 | gagaaa | SC, MC |
| 24b | 1a2 | chr6:22615075-22615078 | caca | SC, MC |
| 24c | 1a2 | chr6:22615080-22615085 | tccaaa | SC, MC |
| 25 | 1a2 | chr6:22615089-22615094 | atgttc | SC, MC |
| 26 | 1a2 | chr6:22615110-22615116 | aaccaca | MC |
| 27 | 1a2 | chr6:22615130-22615136 | gatgag | MC |
| 28 | 1a2 | chr6:22615149-22615155 | cataagg | MC |
| 29 | 1a2 | chr6:22615170-22615180 | agtggcagggt | MC |
| 30 | 1b | chr6:22616161-22616166 | aggaca | MC |
| 31 | 1b | chr6:22616192-22616197 | aggtaa | SC, MC |
| 32* | 1b | chr6:22616210-22616215 | tgggat | SC, MC |
| 33 | 1b | chr6:22616224-22616229 | gatacg | MC |
| 34 | 1b | chr6:22616248-22616256 | gtttctttt | SC, MC |
| 35 | 1b | chr6:22616311-22616316 | gtggat | SC, MC, TM4 |
| 36 | 1b | chr6:22616328-22616340 | ttaattagtataa | MC, TM4, MSC1 |
| 37 | 1b | chr6:22616360-22616370 | gagtaca | SC, MC |

** Indicates different footprint pattern between cell types*

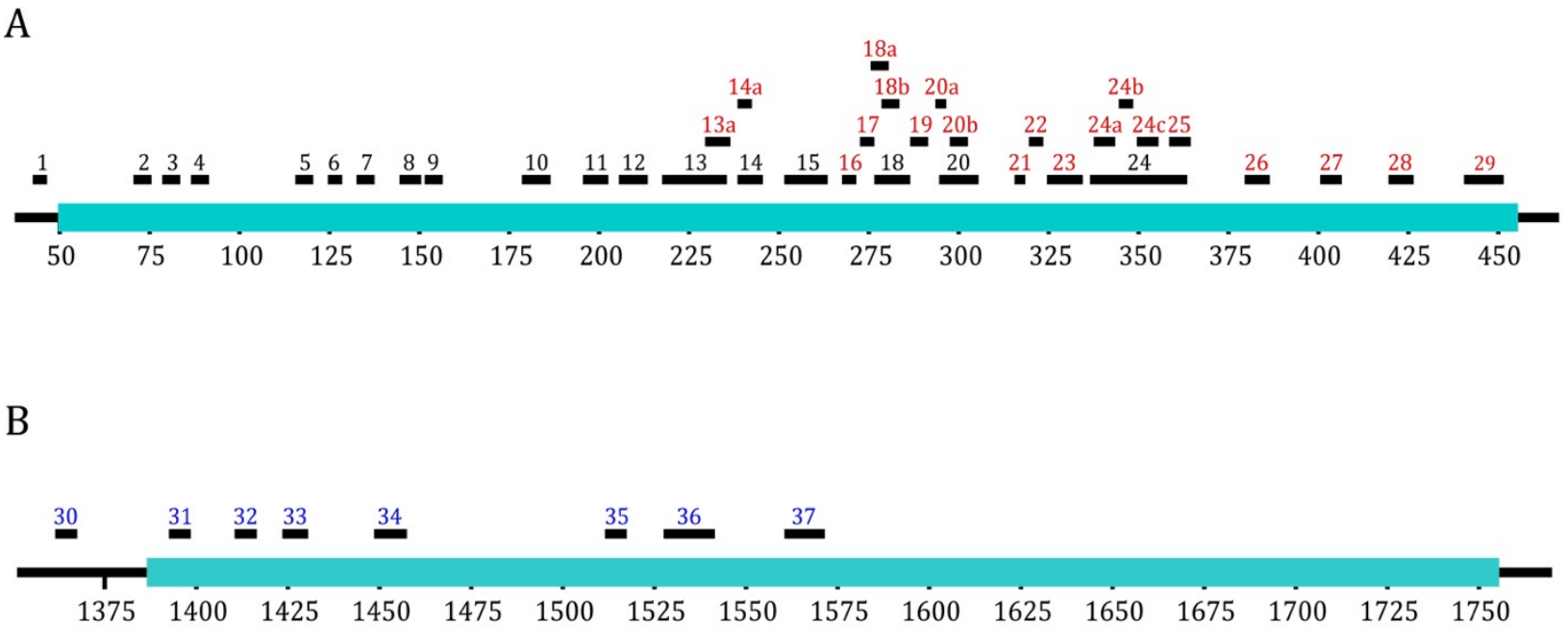



**Figure 4-11. Annotated DNase I footprints – ECR 1f-a and ECR 1f-b.** Results from DNaseI footprinting analysis annotated on (A) ECR 1f-a and (B) ECR 1f-b. The light blue bars on the sequence represent ECR's. The black horizontal bars above the sequence identify footprints their corresponding number is located above. Font colors above the footprint represent the corresponding probe. Black identifies probe 1a1, Red identifies probe 1a2 and blue identifies probe 1b. The ruler below the sequence line represents size in base pairs.

**ECR1f-a and ECR1f-b block deletions identify single region capable of relieving promoter repression.** To identify the regulatory sequences within ECR 1f, various regions within ECR 1f-a and ECR 1f-b were systematically deleted from the full-length ECR 1f and assayed for activity by transient transfections in Sertoli and myoid cells (Figure 4-12 A). Myoid cell data is not shown, as the standard deviation was too large to obtain meaningful data. Within Sertoli cells, deletions in regions 1-4 correspond to ECR 1f-a and deletions in regions 5-7 correspond to ECR 1f-b. Each deletion construct was cloned into the FSHR -220/+123 plasmid in a 5'-forward orientation.

Transient transfection analysis demonstrates that deletion of regions 1, 2, 5, 6 and 7 cannot relieve the promoter repression induced by ECR 1f (Figure 4-12 B). However, deletion of the entire 1f-a region (region 4) or deletion of its corresponding 3' end (region 3) not only relieved the repression generated by the region, but moderately enhanced promoter activity. These results indicate that the activity of ECR 1f lies in region 3 at the 3' end of ECR 1f-a. This data conflicts with the transient transfection data obtained by cloning of 1f-a in isolation with the *Fshr* promoter. This implies that other regions within ECR 1f are required for full regulation of the *Fshr* promoter.

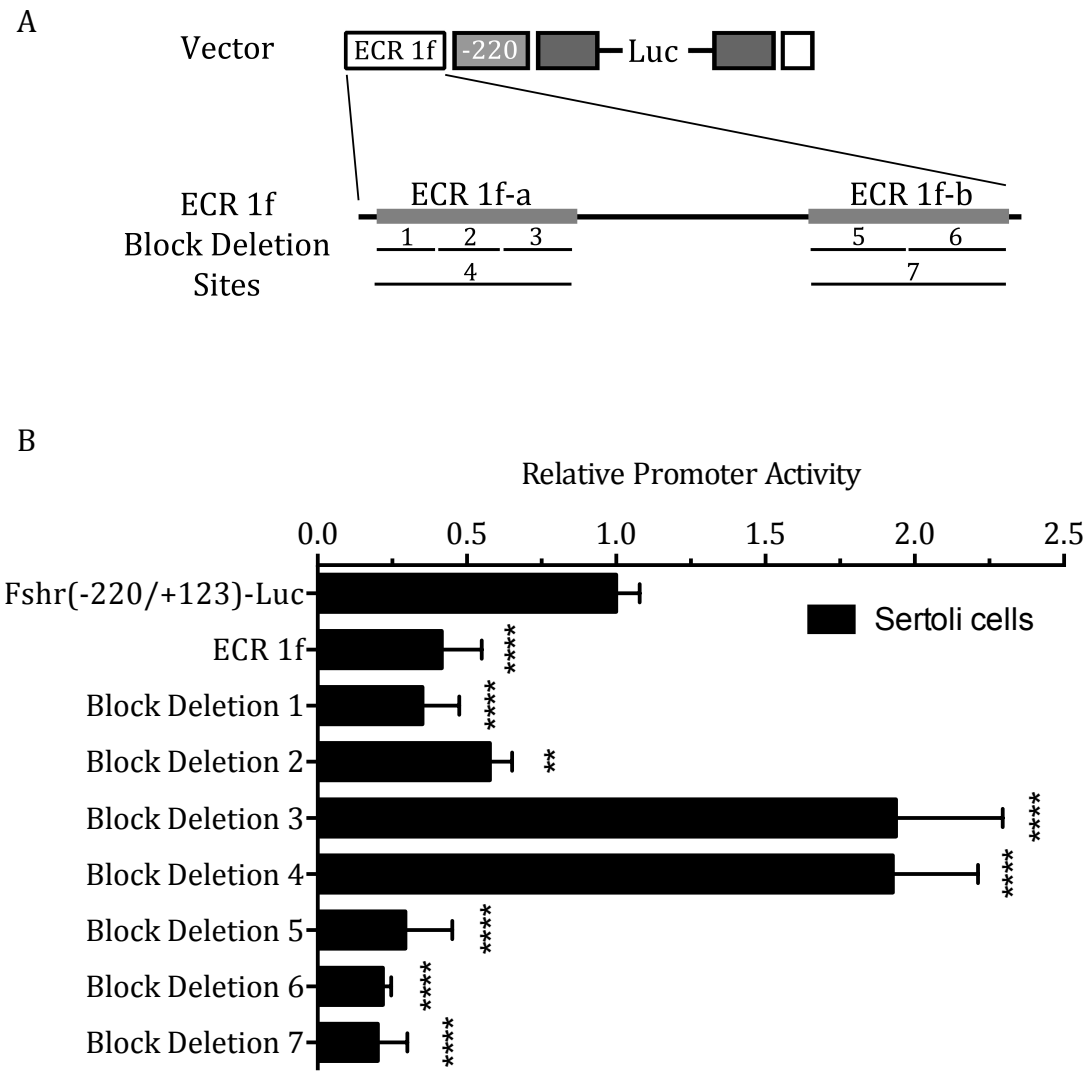



**Figure 4-12. Block deletion of ECR 1f-a region 3 and 4 relieves repression on the *Fshr* promoter.** (A) -200/+123 bp *Fshr* promoter with ECR 1f identifying the location of block deletions generated across ECR 1f-a and 1f-b. (B) Plasmids containing block deletions across ECR 1f-a and ECR 1f-b within ECR 1f were cloned into the pGem vector containing the -200/+123 bp *Fshr* promoter and transiently transfected into Sertoli cells in the 5' forward direction. Error bars reflect the standard deviation, and statistically significant differences are indicated by ** p-value≤0.01 and **** p-value≤0.0001, as determined by one-way ANOVA.

**Individual mutations of footprinted sequences failed to alter ECR 1f activity.** As determined by the DNaseI footprinting experiments, nine sites (FP 21-29) lie within ECR 1f-a region 3 (Figure 4-13, a). To determine if any of these regions are associated with *Fshr* promoter function *in vitro*, we selectively mutated each footprint region within ECR 1f and tested each construct in transient transfections. Bases within the footprinted regions were randomly changed for analysis. For these transfections, MSC1 cells (a mouse Sertoli cell line) were used instead of Sertoli cells to increase the speed and efficiency of construct analysis. ECR 1f represses the *Fshr* promoter within this cell-type ~2.5 fold. Within this system, none of the mutations generated relieved the repressive effects of ECR 1f on the *Fshr* promoter (Figure 4-13, b). Each mutation was made individually and combinations of the mutations were not tested.

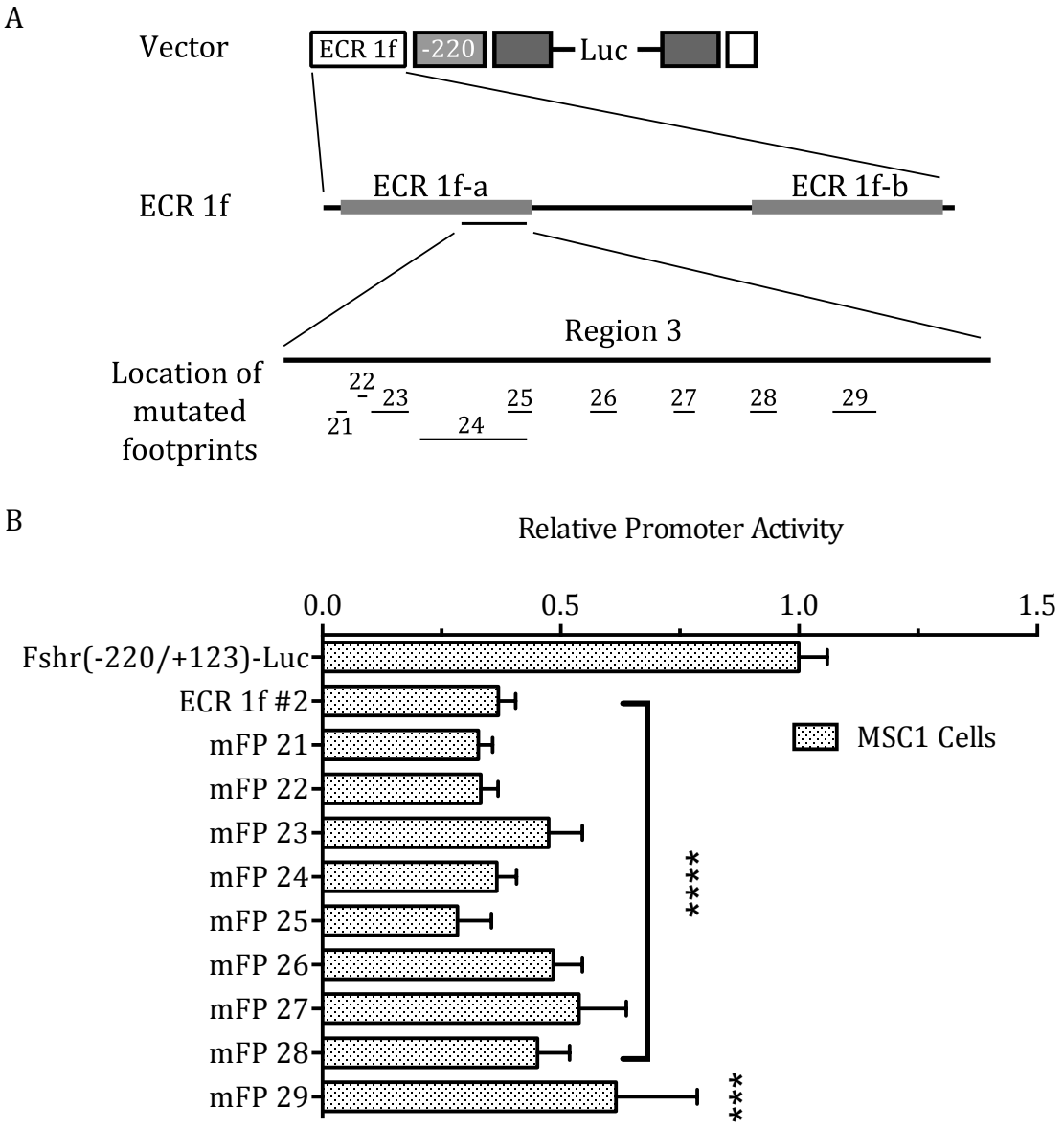



**Figure 4-13. Identified footprint regions do not individually affect *Fshr* promoter function *in vitro*.** (A) -200/+123 bp *Fshr* promoter with ECR 1f identifying the location footprint mutations generated across ECR 1f-a Region 3. (B) plasmids containing mutations across each specific footprint were cloned into the pGem vector upstream of the -200/+123 bp *Fshr* promoter and transiently transfected into MSC1 cells. Error bars reflect the standard deviation, and statistically significant differences are indicated by *** p-value≤0.001 and **** p-value≤0.0001, as determined by one-way ANOVA.

**Block deletions within ECR 1f-a region 3 refined regions of activity**. Since mutation of the footprinted sequences in ECR 1f-a Region 3 did not identify any active sequences in this region, smaller 40 bp deletions were generated across region 3 and assayed for activity in MSC1 cells (Figure 4-14 A). Similar to the transfection in Sertoli cells, deletion of Region 3 relieved the repressive effect of ECR 1f on the *Fshr* promoter; however, the effect is not as strong as previously seen in the primary Sertoli cell transfections. Block deletions across regions 3.1, 3.2, 3.3 and 3.4 were not significantly different from ECR 1f. A combination of block deletions 3.2 and 3.3 (deletion 3.8 figure 4-14, b) was the smallest regions associated with a relief of repression similar to that of Region 3 in the MSC1 cells. This indicates that the active sequences may reside in this region.

The promoter activity observed by deletion of ECR 1f region 3 in the MSC1 cells does not match that of primary Sertoli cells indicating that the regulatory mechanisms may differ between these two cell types. These deletions also need to be analyzed in primary Sertoli cells to identify the activity of these constructs within the primary cells.

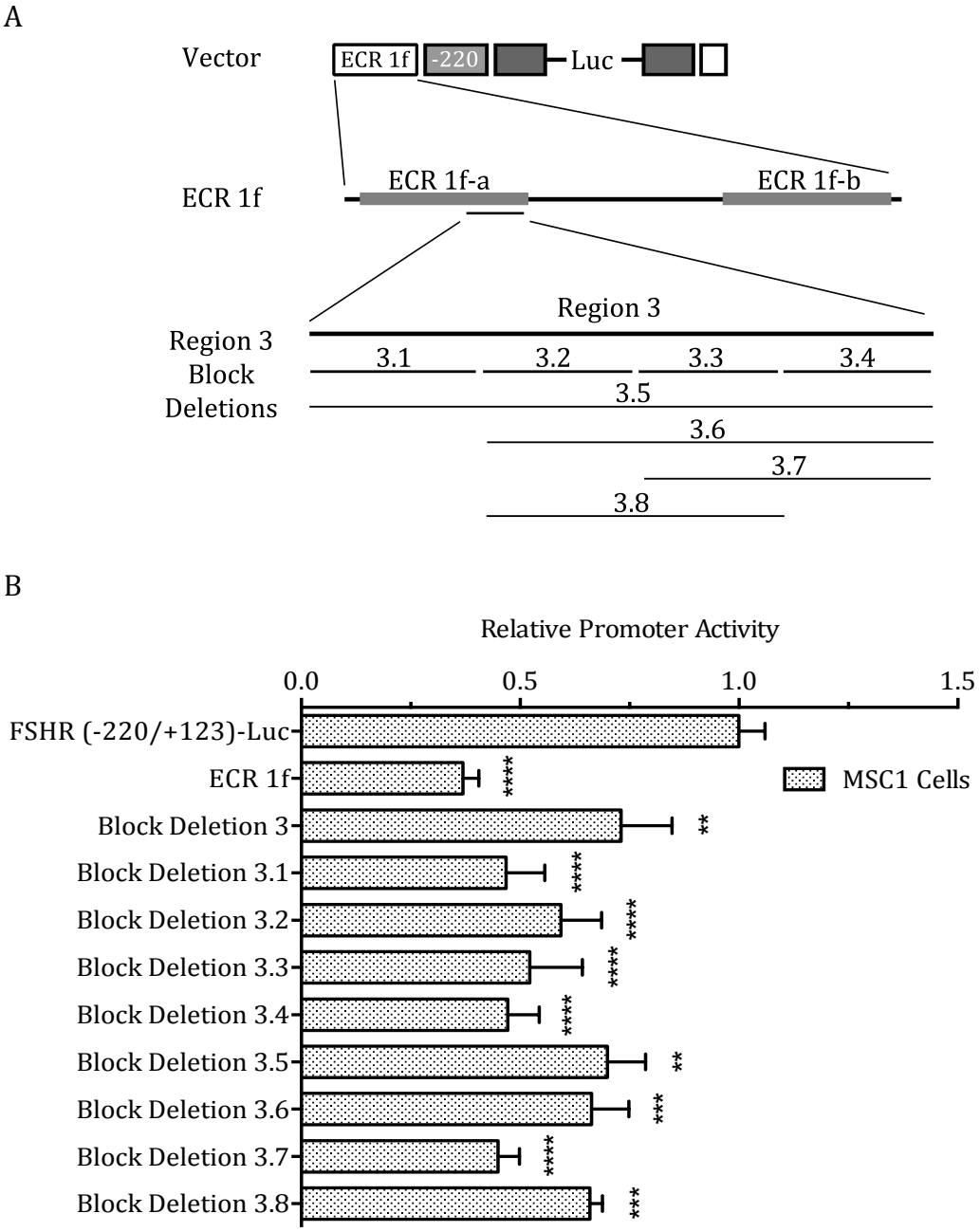



**Figure 4-14. Block deletions across ECR 1f-a region 3.** (A) Annotation of block deletions generated across ECR 1f-a region 3 and (B) plasmids containing block deletions across ECR 1f-a region 3 were cloned into the pGem vector containing the -200/+123 bp *Fshr* promoter and transiently transfected into MSC1 cells. Error bars reflect the standard deviation, and statistically significant differences are indicated by ** p-value≤0.01, *** p-value≤0.001 and **** p-value≤0.0001, as determined by one-way ANOVA.

**EMSA analysis of ECR 1f-a region 3.8 identified a GATA4 binding site**. Since ECR 1f contains a significant portion of the region 3 activity, Electric Mobility Shift Assays (EMSA) were performed to characterize the protein binding regions. Four probes spanning Region 3.8, were used with Sertoli cell nuclear extract (Figure 4-15). No specific complexes were observed on probes 3.2-2 and 3.3-1; However, complexes were observed on 3.2-1 and 3.3-2 (Figure 4-16). Probe 3.2-1 contains three possible GATA binding sites (Figure 4-17), GATA antibodies were added to EMSA reactions to determine their ability to cross-react with bound proteins. Figure 4-18 shows that only the GATA4 antibody is capable of cross-reacting with the observed protein/DNA band. To identify the specific GATA binding site, competitors with mutations in each GATA site were included in the EMSA reactions. Competitors with mutations generated in the first (Mut1 and Mut2) and third (Mut4) GATA binding site competed off almost all the protein/probe interaction indicating that these sites are not required for binding (Figure 4-19). Mutation of the second GATA binding site (Mut3) could not compete off the protein/probe interaction indicating that the second GATA site is responsible for the protein/DNA interaction. Competition with a sequence containing two GATA consensus sequences also fully competed all binding from the probe, further supporting the identification of GATA as the binding protein.

To further characterize proteins binding to probe 3.3-2, six base pair mutations were made across the length of the probe and used as competitors. Two mutant competitors (Mut4 and Mut5) did not compete the protein complexes bound to the probe (Figure 4-20), identifying the sequence 5'- ATT ATT CAT AAG-3' as

critical for protein binding. Comparison of the sequence to a database of known transcription factor binding sites identified several binding candidates, most of which belong to the POU family of transcription factors. Specifically POU1F1a, a.k.a pit-1 and GHF-1, has a similar consensus sequence 5'-TATNCAT-3' (Figure 4-21).

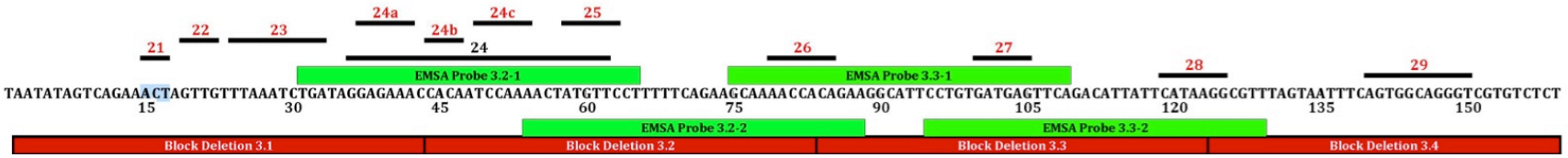



**Figure 4-15. EMSA Probes for ECR 1f-a Region 3.8.** Four probes were generated to span ECR 1f-a region 3.8 and are identified by the green horizontal bars. The red horizontal bars identify genomic location of block deletions.

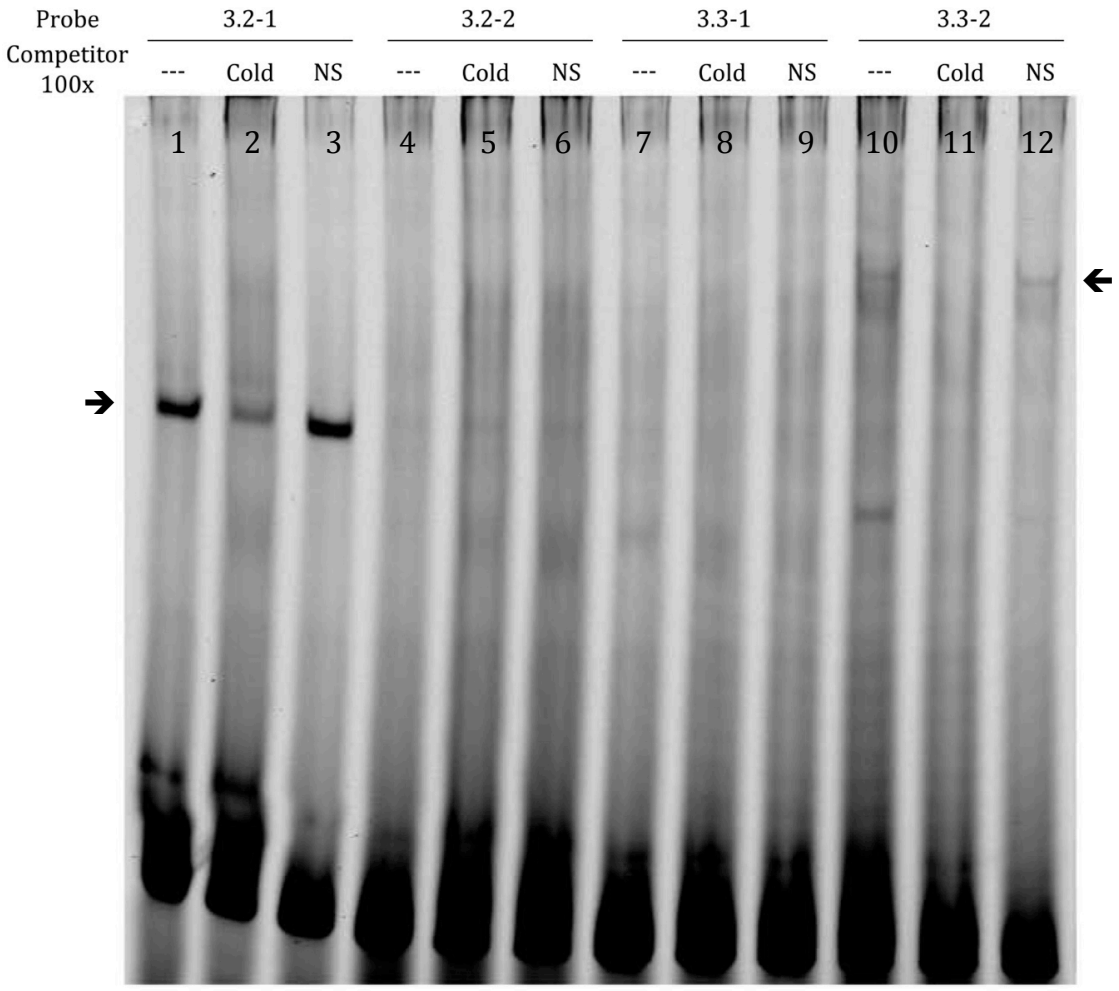



**Figure 4-16. Two regions are capable of binding Sertoli cell nuclear proteins *in vitro.*** EMSA binding assays of four probes spanning ECR 1f Region 3.8. Each probe was incubated with Nuclear extract only (Lanes 1, 4, 7 and 10), unlabeled competitor + nuclear extract (Lanes 2, 5, 8 and 11) and nonspecific competitor + nuclear extract (Lanes 3, 6, 9 and 12). Black arrows indicate shift location of specific protein/DNA interactions.

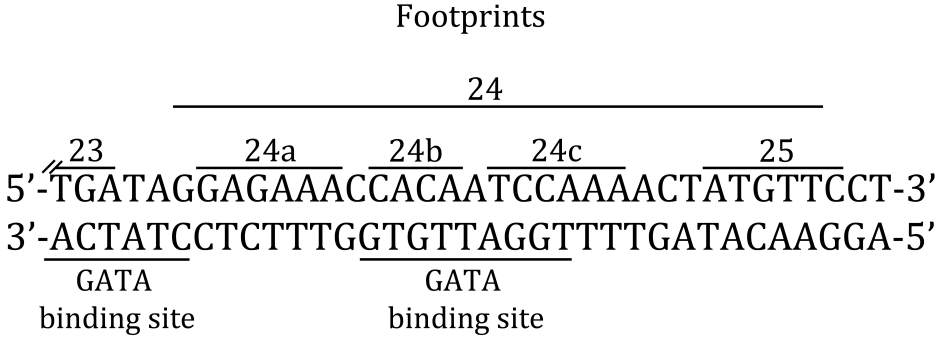



**Figure 4-17. Footprints and potential GATA binding sites within EMSA probe 3.2-1.** EMSA probe 3.2-1 identifying potential GATA binding sites (underlined sequence). Footprints are labeled with above the probe sequence.

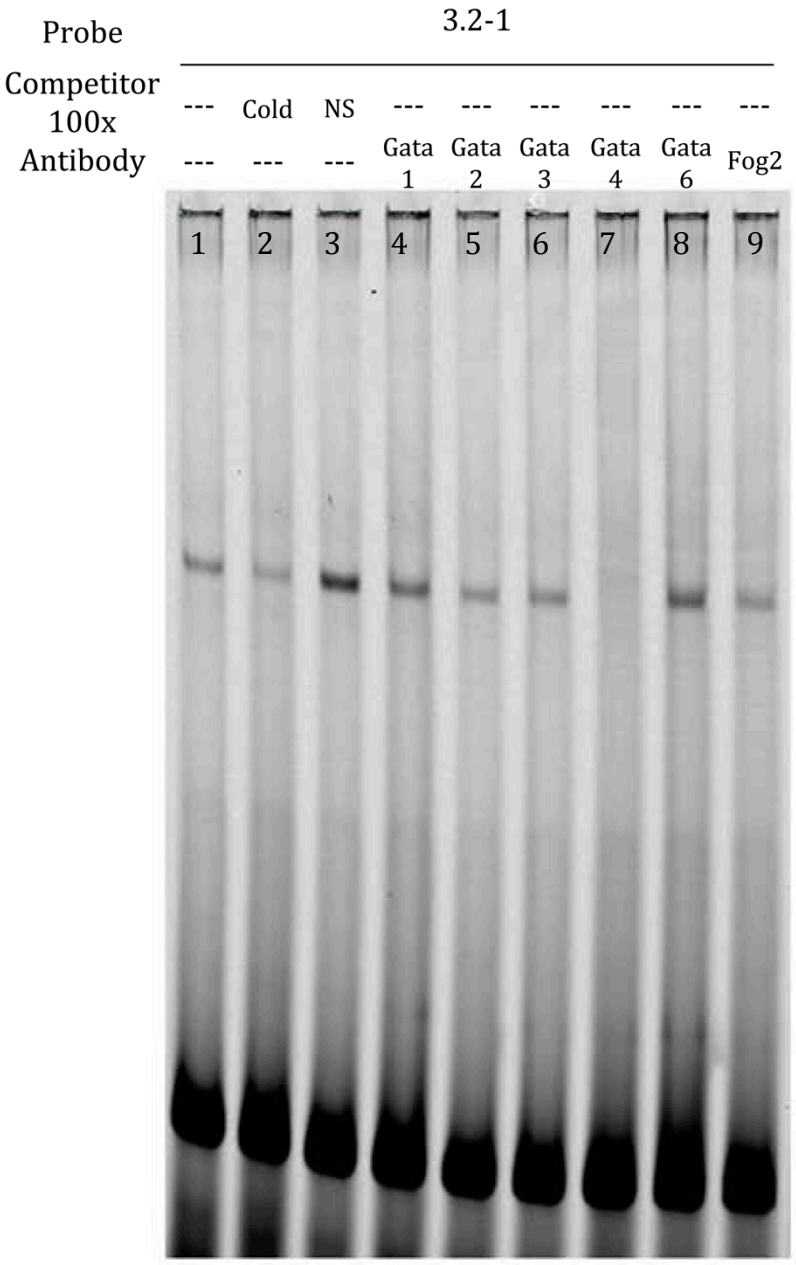



**Figure 4-18. Identification of GATA-4 specific protein/DNA interaction.** EMSA binding assay of probe 3.2-1. Probe was incubated with Nuclear extract only (Lane 1), unlabeled competitor + nuclear extract (Lane 2), nonspecific competitor + nuclear extract (Lane 3) and nuclear extract + GATA specific antibodies (Lanes 4-8) and FOG2 (Lane 9). Only the GATA-4 Antibody was capable of shifting the Probe/protein band (Lane 7).

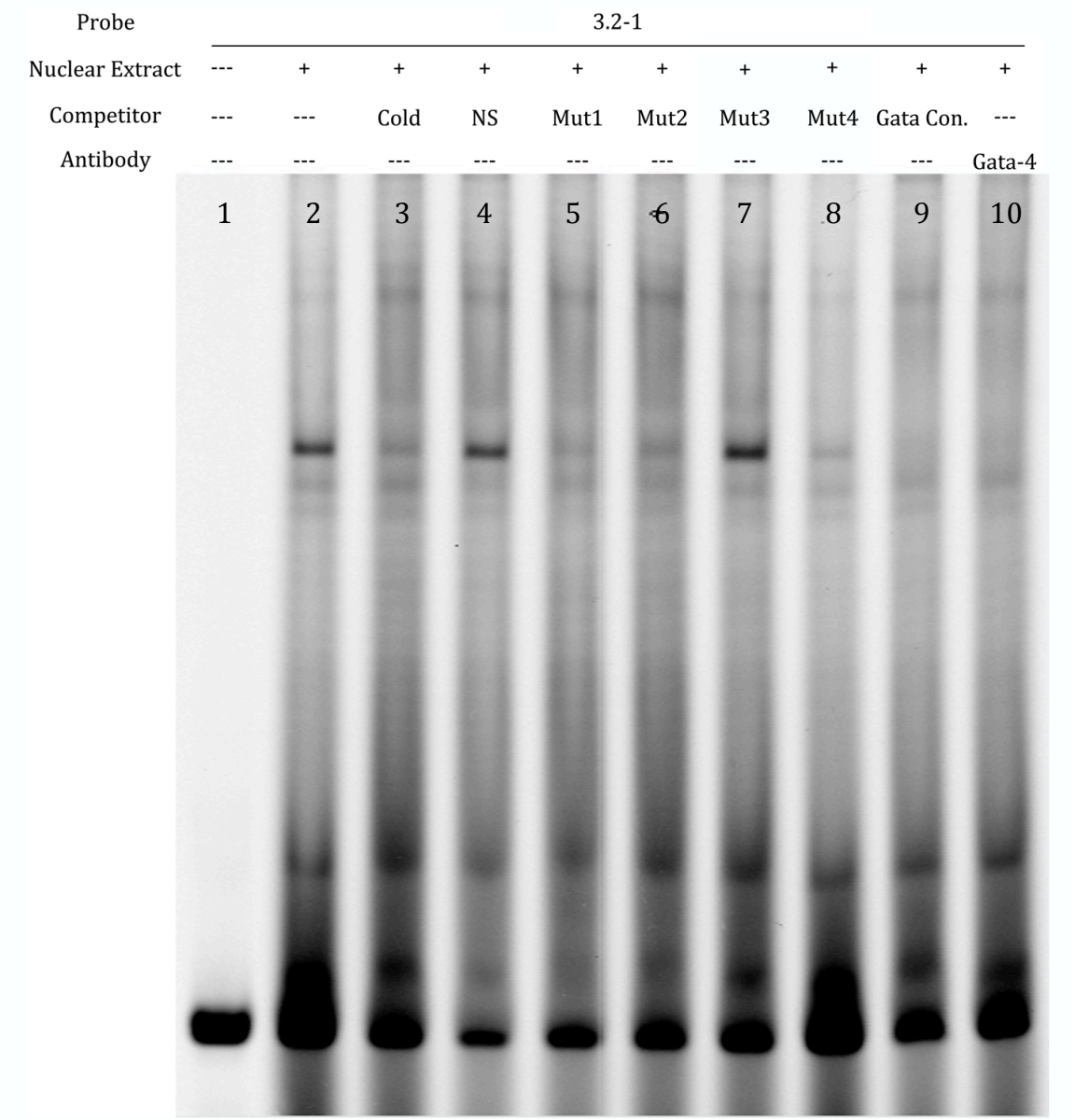



**Figure 4-19. Identification of GATA-4 specific DNA binding site.** EMSA binding assay of probe 3.2-1. Probe was incubated without Nuclear extract (Lane 1), Probe was incubated with Nuclear extract (Lane 2), unlabeled competitor + nuclear extract (Lane 3), nonspecific competitor + nuclear extract (Lane 4) and nuclear extract + GATA site specific mutations (Lanes 5-8), GATA consensus sequence (Lane 9) and GATA-4 antibody (Lane 10).

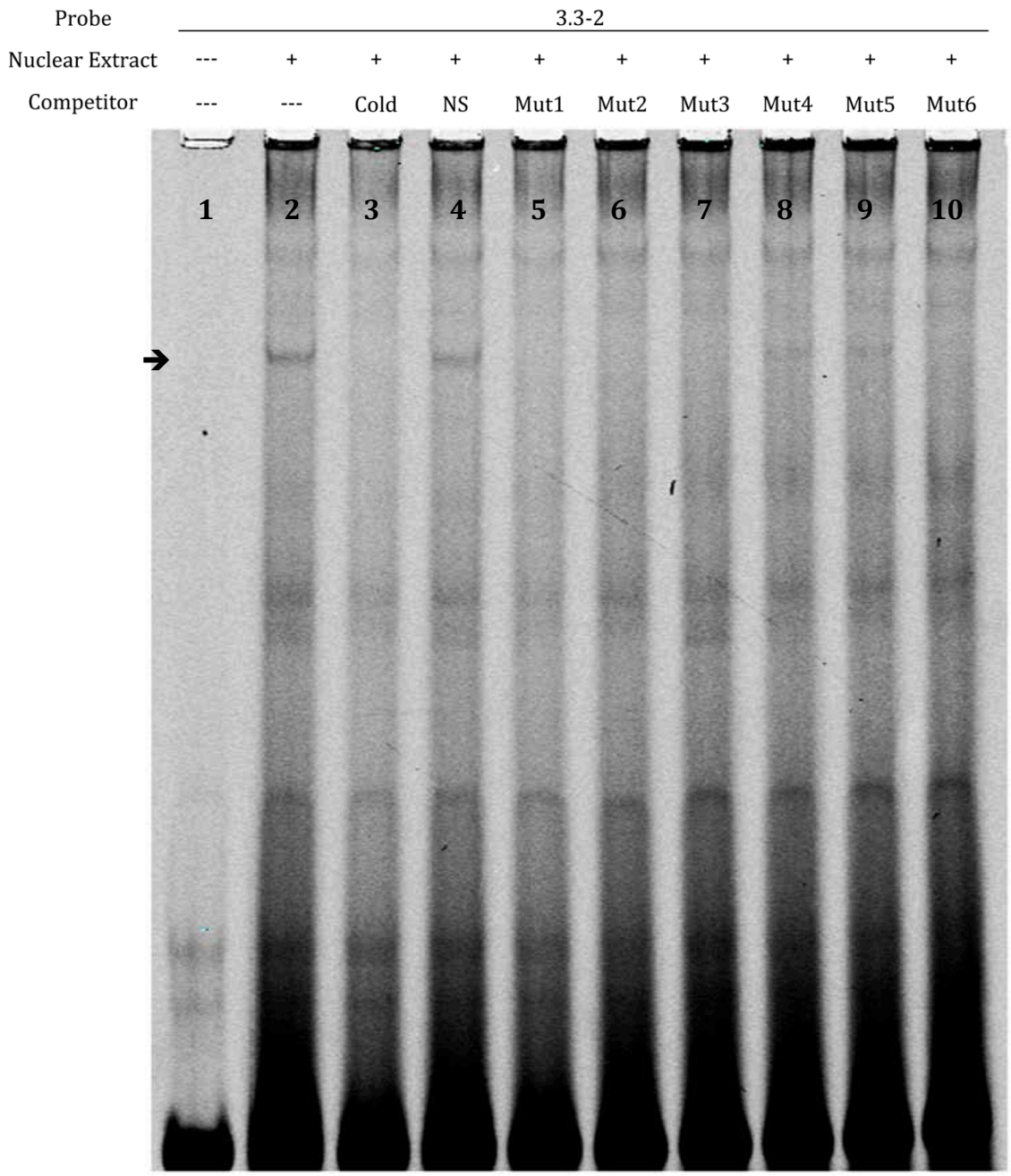



**Figure 4-20. Identification of 12 base pair binding sequence within EMSA probe 3.3-2.** EMSA binding assay of probe 3.3-2. Probe was incubated without Nuclear extract (Lane 1), nuclear extract (Lane 2), unlabeled competitor + nuclear extract (Lane 3), nonspecific competitor + nuclear extract (Lane 4) and competitors containing 6 bp mutations across the entire probe (Lanes 5-10). Only mutant competitors 4 and 5 are incapable of competing the probe/protein interaction.

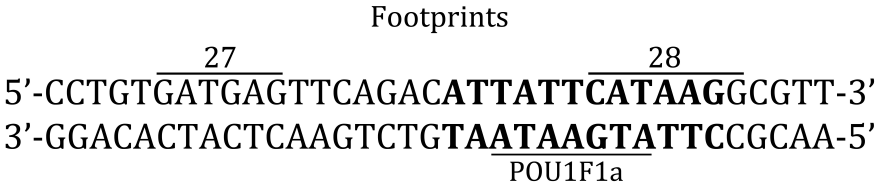



**Figure 4-21. EMSA probe 3.3-2 identifying potential POU1f1 binding site.**
EMSA probe 3.3-2 identifying potential POU1fa1 binding site (underlined sequence). Bold font indicates sequence identified by mutated competitors 4 and 5. Footprints are labeled with above the probe sequence.

# Discussion

To date, most research on the regulation of *Fshr* has focused on identifying regulatory elements located in its 5' flanking region, making this region the most highly characterized of the locus. *In vitro* transient transfection assays of the wild type and mutant promoters identified elements required for activity of the 5' flanking region including an E-box, AP1, E2F, Inverted GATA and Inr regulatory sites (Goetz, Lloyd et al. 1996; Heckert, Daggett et al. 1998; Heckert, Sawadogo et al. 2000; Kim and Griswold 2001). This characterization identified an E-box element just upstream of the transcriptional start site, as the dominant regulatory element in this region. *In vitro* and *in vivo* analysis identified the transcription factors Upstream Regulatory Factor 1 and 2 (USF1 and USF2) as the main E-box bound proteins (Goetz, Lloyd et al. 1996; Heckert, Daggett et al. 1998). *In vivo* footprinting and ChIP analysis for USF1/2 indicate that E-box is only occupied in *Fshr*-expressing cells (Heckert, Sawadogo et al. 2000; Hermann, Hornbaker et al. 2008). Therefore, it appears specificity is somehow linked to access of USF1/2 to the promoter. USF1 and USF2 proteins are ubiquitously expressed transcription factors and have been associated with CBP/p300 and recruitment of proteins associated with histone modifications (Sirito, Lin et al. 1994; Breen and Jordan 2000). Although USF proteins are required for *Fshr* transcriptional activity, their ubiquitous expression makes them unlikely determinants of *Fshr* cell-specificity, which is limited to Sertoli and granulosa cells. Thus, other mechanisms or elements must be required to enable USF proteins to have access to the promoter and initiate gene transcription.

To elucidate which sequences required for correct spatio-temporal expression of *Fshr*, transgenic animals were created with various amounts of the *Fshr* promoter directing a LacZ reporter (Linder, Heckert et al. 1994; Heckert, Sawadogo et al. 2000; Hermann and Heckert 2007). Analysis of the animals confirmed the requirement for distal elements as none of the transgenes expressed appropriately, even for YAC transgenes containing up to 413 kilobases of the entire locus(Hermann, Hornbaker et al. 2007). Taken together, this data suggests that correct spatio-temporal expression of *Fshr* requires regulatory elements that lie far from the promoter.

Identification of distal regulatory elements has not only increased our understanding of gene regulation, but has also led to the identification of numerous distal elements important in development and disease (Koschmieder, Rosenbauer et al. 2005; Rijnkels, Kabotyanski et al. 2010; Hardison 2012; Peterson, Fedosyuk et al. 2012; Xu, Shao et al. 2012). Identifying distal regulatory elements has become easier due to the large amount of whole genome sequence data now available. Considerable data from the literature indicate that sequence conservation is a hallmark of functional importance and this is especially true for regulatory elements (Dickmeis and Muller 2005; Hawkins and Ren 2006; Acevedo, Iniguez et al. 2007; Visel, Bristow et al. 2007; Brown 2008; Strahle and Rastegar 2008; Fromm and Bulger 2009). As noted in the results, we mapped the conservation profile and 7x regulatory potential of the 1.4 Mb locus of the *Fshr* and identified 31 ECRs associated with increased regulatory potential. While computational genomics has helped identify new distal elements, false positives and undetected sequences limit

the analysis (Birney, Stamatoyannopoulos et al. 2007; Giresi, Kim et al. 2007; Margulies, Cooper et al. 2007).

To overcome this limitation, our bioinformatic selection was further refined using ChIP-qPCR and ChIP-Seq analysis of histone modifications across the *Fshr* locus. Numerous reports now exist that use histone modifications to identify important regulatory elements (Rye, Saetrom et al. 2011; Shu, Chen et al. 2011; Choe, Hong et al. 2012; Zentner and Scacheri 2012). For our studies, a genome-wide map of H3K4me3, a chromatin modification known to associate with regulatory elements, was generated for both Sertoli and myoid cells. These maps identified a single conserved region associated with cell-type specific H3K4me3 enrichment– ECR 1f.

ChIP-qPCR analysis was used to gain a more in-depth understanding of the chromatin environment encompassing ECR 1f. The histone modifications enriched within ECR 1f include H4ac, H3K4me3 and to a lesser extent H3ac. Each of these marks is typically associated with active chromatin regions; further indicating that ECR 1f is actively involved in regulation within Sertoli cells. Enrichment of histones corresponding to repressed regions, H3K27me3 and H3K9me2 was not detected. These assays demonstrate that the chromatin surrounding ECR 1f is permissive to activity.

To identify the functional activity of ECR 1f with the *Fshr* promoter, vector constructs were transiently transfected into Sertoli cells. Transient transfection analysis revealed that ECR 1f could repress the *Fshr* promoter activity *in vitro* in both primary Sertoli cell and myoid cultures. The repressive activity of ECR 1f was

not restricted by orientation within the plasmid and did not change when different promoter sizes (-220 or -2700 bp) were used indicating that ECR 1f is an insulator. The repressive activity observed was also not restricted by cell type, indicating that the transcription factors responsible for the repressive activity are present in both Sertoli and myoid cells.

Perhaps the most surprising finding was that ECR 1f repressed *Fshr* promoter activity *in vitro.* Data obtained from our ChIP studies indicated that ECR 1f was enriched for histone modifications commonly associated with activity. The most likely explanation for this is that not all distal regulatory elements function the same *in vitro* as they do *in vivo* (Bulger and Groudine 2010). When regulatory elements are taken out of their native chromatin environment and placed in reporter assays, several factors are altered including the lack of appropriate histone association, limiting physical constraints and the possible removal of other key regulatory sequences involved. Another possibility is that ECR 1f is functioning to actively repress the *Fshr in vivo*. However, a search of the literature did not identify other distal regulatory elements associated with an active histone profile that in turn repressed gene activity *in vivo.*

To further explore the regulatory sequence within ECR 1f, *in silico* was used to identify highly conserved regions within ECR 1f. Two regions were identified, ECR 1F-a and ECR 1F-b, and assayed for activity using *Fshr* promoter constructs. Neither ECR 1f-a or 1f-b was capable of repressing *Fshr* promoter activity to the same level as ECR 1f. However, ECR 1f-b did repress promoter activity in a position and orientation dependent manner and ECR 1f-a activated promoter activity in a

cell-type, position and orientation specific manner. These results indicate that the full length ECR is required for full activity. A fusion construct consisting of ECR 1f-a and ECR 1f-b did repress *Fshr* promoter activity, although not to the same level as the full length ECR 1f implying that the less conserved sequence may play a role in the activity of ECR 1f, either by providing spatial distance between ECR 1f-a and 1f-b or by the addition of other regulatory sequences.

Overall these results imply several possible regulatory actions for ECR 1f. First, it is likely that there are several regulatory elements residing in ECR 1f as indicated by the mixed results obtained using ECR 1f-a and 1f-b. Second, the lack of activity associated with a majority of the ECR 1f-a and ECR 1f-b constructs tested implies that multiple sites are required for full activity. Third, the less conserved region residing in between ECR 1f-a and 1f-b likely plays a role in the regulatory action of ECR 1f. These results indicate that ECR 1f is complex and that multiple regions are required for proper function.

To assess the regulatory potential of individual regions of ECR 1f within the context of the full length ECR 1f, small block deletions were made across ECR 1f-a and ECR 1f-b. By only deleting a subset of each region and leaving the remaining ECR intact, we were able to identify a small region (~150 bp) within ECR 1f-a that, when removed, relieved the repressive effects of ECR 1f on the *Fshr* promoter in Sertoli cells, identified as ECR 1f-a region 3. To further refine this region, 40 base pair block deletions were made across ECR 1f-a region 3 and analyzed for activity in MSC1 cells. This selection process narrowed down the repressive region to ~80 base pairs. *In vitro* footprinting identified 5 possible binding sites within this region,

however individual mutation analysis of these 5 regions was unable to relieve the repressive effect of ECR 1f on the *Fshr* promoter.

*In silico* analysis of transcription factor-binding sites within ECR 1f-a region 3 identified two possible regulatory candidates- GATA and POU. GATA-4 is a transcription factor whose cell specific expression is limited to the heart, testis and ovary. Within the testis, GATA-4 expression is restricted to Sertoli and Leydig cells. Expression has been documented from embryonic day 11.5 and remains throughout adulthood; however, the highest expression levels are found in Sertoli cells during sex determination. EMSA analysis selectively identified GATA-4 as the GATA family member bound to ECR 1f-a region 3 *in* vitro.

Previous studies from our laboratory identified a repressive site located within the first intron of *Fshr* that also binds GATA-4 *in vitro* (Hermann and Heckert 2005). In the current study, we identified another distal region that is capable of repressing *Fshr* promoter activity, *in vitro*, in Sertoli and myoid cells alike and found that it is capable of binding GATA-4 *in vitro*. ECR 1f lies ~400 kb downstream of the *Fshr* promoter, indicating that distal looping must occur to enable this region to regulate the promoter. Studies have indicated that other members of the GATA family are capable of achieving distal interactions and can act as either activators or repressors depending on gene context (Fong and Emerson 1992; Barton, Madani et al. 1997; Grass, Boyer et al. 2003; Obara, Suzuki et al. 2008; Yu, Riva et al. 2009; Chen, Bates et al. 2012). It is possible that the two GATA sites identified work in concert together to regulate the *Fshr* promoter in Sertoli cells.

*In silico* analysis also identified a consensus sequence in ECR 1f-a region 3 compatible with the POU family of transcription factors, specifically POU1f1, a.k.a. pit-1. The primary role of POU1f1 is regulating pituitary development and hormone secretion, however, a splice variant of pit-1, pit-1w was identified in the testis and its role is currently under investigation (Taniuchi, Maeda et al. 2011; Maeda, Taniuchi et al. 2012). The role of this protein was not explored in the current study, as antibodies have not been identified to use in EMSA analysis. However, mutant competitors did identify a 12 base pair sequence capable of binding protein from Sertoli cell nuclear extract and this 12 base pair sequence matches the POU consensus sequence.

The current study expands our knowledge of the regulatory landscape surrounding *Fshr in vivo* and has identified two more sites capable of repressing the *Fshr* promoter *in vitro*. Future studies will aim to decipher the role of this distal element *in vivo.*

# Chapter 5

# Future Directions

## ChIP-Seq Analysis of Histone Modifications

Our initial objective was to map genome wide histone modifications in P15Sprague Dawley Sertoli and peritubular myoid cells for the use of identifying Sertoli cell specific distal regulatory elements. For both of these cell types, we gathered enrichment data for H4ac, H3K4me3 and H3K27me3. The raw sequence tags were processed and aligned to the latest version of the *Rattus norvegicus* genome (rn5, March 2012). Each aligned sequence file was processed for viewing on the UCSC genome browser. Unfortunately, the alignment quality for the Sertoli cell samples H4ac and H3K27me3 were poor and meaningful data could not be extracted from the alignments. For complete analysis to be performed between the two cell types, these two samples will need re-sequenced.

The H3K4me3 samples for both Sertoli and myoid cells did align well to the reference genome and provided good quality data. For this dissertation, only the top 40 enriched regions were considered for analysis. Although meaningful, this analysis only contained a fraction of the enriched regions identified. Future work with these datasets should include analysis across all of the identified regions, which will lead to a complete picture of the H3K4me3 landscape within Sertoli and myoid cells. Future analysis should also include RNA-Seq analysis identifying all the transcribed sequences within Sertoli and myoid cells at P15.

Completion of this work will generate a complete picture of the histone environment for H4ac, H3K4me3 and H3K27me3. Combining this data with cell-specific expression analysis will greatly aid in the identification of cell-specific regulatory sequences. This dataset will be a great asset to the understanding of

transcriptional regulation in both Sertoli and myoid cells and will allow for the identification of Sertoli cell specific modification patterns. This dataset will also provide a comprehensive map that can be used in the identification of new regulatory elements important in Sertoli cell transcriptional regulation.

## Distal Regulation of *Fshr*

Using a combination of sequence conservation and H3K4me3 profiles across the *Fshr* locus we identified a putative *cis*-regulatory element (ECR 1f) at the 3' end of *Fshr*. Initial *in vitro* analysis indicates that ECR 1f represses the *Fshr* promoter in Sertoli cells and that GATA-4 may be responsible for this activity. *In vitro* protein/DNA binding analysis also identified another site capable of binding a Sertoli cell nuclear protein that may be involved and *in silico* analysis indicates the likely candidate belongs to the POU family of transcription factors. Further EMSA analysis with POU1f1 specific antibodies could confirm this finding.

Aside from the *in vitro* binding studies, future studies will include *in vivo* analysis of GATA-4 binding within ECR 1f using chromatin immunoprecipitation. To determine if ECR 1f is in close physical contact with the *Fshr* promoter, Chromatin conformation capture (3C) will be performed. This technique allows for the identification of distal regulatory regions that are in close physical contact with known promoter regions *in vivo*. Identification of the distal element regulating *Fshr* will further our understanding of transcriptional regulation in Sertoli cells and provide information regarding the regulation of *Fshr*. Identification of the transcription factors involved in the regulation of *Fshr* will further our understanding of transcriptional regulation in Sertoli cells. These results are not

only relevant for Sertoli cells as *Fshr* is also expressed in ovarian granulosa cells. Understanding the mechanisms and regulatory proteins involved in *Fshr* expression will further our working knowledge and understanding of the hypothalamic-pituitary-gonadotropin axis.

# Chapter 6

# References

Abou-Issa, H. and L. E. Reichert, Jr. (1977). "Solubilization and some characteristics of the follitropin receptor from calf testis." J Biol Chem **252**(12): 4166-4174.

Acevedo, L. G., A. L. Iniguez, et al. (2007). "Genome-scale ChIP-chip analysis using 10,000 human cells." Biotechniques **43**(6): 791-797.

Achermann, J. C. and J. L. Jameson (1999). "Fertility and infertility: genetic contributions from the hypothalamic-pituitary-gonadal axis." Mol Endocrinol **13**(6): 812-818.

Akalin, A., D. Fredman, et al. (2009). "Transcriptional features of genomic regulatory blocks." Genome Biol **10**(4): R38.

Akkers, R. C., S. J. van Heeringen, et al. (2009). "A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos." Dev Cell **17**(3): 425-434.

Amouyal, M. (1991). "The remote control of transcription, DNA looping and DNA compaction." Biochimie **73**(10): 1261-1268.

Aparicio, O., J. V. Geisberg, et al. (2005). "Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo." Curr Protoc Mol Biol **Chapter 21**: Unit 21 23.

Barrett, T., D. B. Troup, et al. (2007). "NCBI GEO: mining tens of millions of expression profiles--database and tools update." Nucleic Acids Res **35**(Database issue): D760-765.

Barrett, T., D. B. Troup, et al. (2009). "NCBI GEO: archive for high-throughput functional genomic data." Nucleic Acids Res **37**(Database issue): D885-890.

Barton, M. C., N. Madani, et al. (1997). "Distal enhancer regulation by promoter derepression in topologically constrained DNA in vitro." Proc Natl Acad Sci U S A **94**(14): 7257-7262.

Becker, J., C. Yau, et al. (2013). "NucleoFinder: a statistical approach for the detection of nucleosome positions." Bioinformatics **29**(6): 711-716.

Bejerano, G., C. B. Lowe, et al. (2006). "A distal enhancer and an ultraconserved exon are derived from a novel retroposon." Nature **441**(7089): 87-90.

Ben-Josef, E., S. Y. Yang, et al. (1999). "Hormone-refractory prostate cancer cells express functional follicle-stimulating hormone receptor (FSHR)." J Urol **161**(3): 970-976.

Benson, B., S. Sorrentino, et al. (1969). "Increase in serum FSH following unilateral ovariectomy in the rat." Endocrinology **84**(2): 369-374.

Bernstein, B. E., M. Kamal, et al. (2005). "Genomic maps and comparative analysis of histone modifications in human and mouse." Cell **120**(2): 169-181.

Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-326.

Birney, E., J. A. Stamatoyannopoulos, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Bishop, C. E., D. J. Whitworth, et al. (2000). "A transgenic insertion upstream of sox9 is associated with dominant XX sex reversal in the mouse." Nat Genet **26**(4): 490-494.

Bogerd, J. (2007). "Ligand-selective determinants in gonadotropin receptors." Mol Cell Endocrinol **260-262**: 144-152.

Boitani, C., M. Stefanini, et al. (1995). "Activin stimulates Sertoli cell proliferation in a defined period of rat testis development." Endocrinology **136**(12): 5438-5444.

Braun, T., P. R. Schofield, et al. (1991). "Amino-terminal leucine-rich repeats in gonadotropin receptors determine hormone selectivity." EMBO J **10**(7): 1885-1890.

Breen, G. A. and E. M. Jordan (2000). "Upstream stimulatory factor 2 stimulates transcription through an initiator element in the mouse cytochrome c oxidase subunit Vb promoter." Biochim Biophys Acta **1517**(1): 119-127.

Bressler, R. S. and M. H. Ross (1973). "On the character of the monolayer outgrowth and the fate of the peritubular myoid cells in cultured mouse testis." Exp Cell Res **78**(2): 295-302.

Brown, C. T. (2008). "Computational approaches to finding and analyzing cis-regulatory elements." Methods Cell Biol **87**: 337-365.

Brune, M., C. Adams, et al. (2010). "Primate FSH-receptor promoter nucleotide sequence heterogeneity affects FSH-receptor transcription." Mol Cell Endocrinol **317**(1-2): 90-98.

Bulger, M. and M. Groudine (2010). "Enhancers: the abundance and function of regulatory sequences beyond promoters." Dev Biol **339**(2): 250-257.

Calestagne-Morelli, A. and J. Ausio (2006). "Long-range histone acetylation: biological significance, structural implications, and mechanisms." Biochem Cell Biol **84**(4): 518-527.

Camp, T. A., J. O. Rahal, et al. (1991). "Cellular-Localization and Hormonal-Regulation of Follicle-Stimulating-Hormone and Luteinizing-Hormone Receptor Messenger-Rnas in the Rat Ovary." Molecular Endocrinology **5**(10): 1405-1417.

Cannon, J. G., M. Cortez-Cooper, et al. (2009). "Follicle-stimulating hormone, interleukin-1 and bone density in adult women." Am J Physiol Regul Integr Comp Physiol.

Chaudhary, J., P. D. Whaley, et al. (1996). "Transcriptional regulation of sertoli cell differentiation by follicle-stimulating hormone at the level of the c-fos and transferrin promoters." Biol Reprod **54**(3): 692-699.

Chauvigne, F., A. Tingaud-Sequeira, et al. (2010). "Functional and Evolutionary Analysis of Flatfish Gonadotropin Receptors Reveals Cladal- and Lineage-Level Divergence of the Teleost Glycoprotein Receptor Family." Biol Reprod.

Chemes, H. E., M. Dym, et al. (1979). "Hormonal regulation of Sertoli cell differentiation." Biol Reprod **21**(1): 251-262.

Chen, H. P., A. Lin, et al. (2008). "Screening reveals conserved and nonconserved transcriptional regulatory elements including an E3/E4 allele-dependent APOE coding region enhancer." Genomics **92**(5): 292-300.

Chen, J. K. and L. L. Heckert (2001). "Dmrt1 expression is regulated by follicle-stimulating hormone and phorbol esters in postnatal Sertoli cells." Endocrinology **142**(3): 1167-1178.

Chen, Y., D. L. Bates, et al. (2012). "DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation." Cell Rep **2**(5): 1197-1206.

Choe, M. K., C. P. Hong, et al. (2012). "Functional elements demarcated by histone modifications in breast cancer cells." Biochem Biophys Res Commun **418**(3): 475-482.

Choi, J. H., K. C. Choi, et al. (2004). "Overexpression of follicle-stimulating hormone receptor activates oncogenic pathways in preneoplastic ovarian surface epithelial cells." J Clin Endocrinol Metab **89**(11): 5508-5516.

Conley, A. B. and I. K. Jordan (2010). "Identification of transcription factor binding sites derived from transposable element sequences using ChIP-seq." Methods Mol Biol **674**: 225-240.

Consortium, I. C. G. S. (2004). "Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution." Nature **432**(7018): 695-716.

Cooper, G. M., M. Brudno, et al. (2004). "Characterization of evolutionary rates and constraints in three Mammalian genomes." Genome Res **14**(4): 539-548.

Cui, P., W. Liu, et al. (2012). "Comparative analyses of H3K4 and H3K27 trimethylations between the mouse cerebrum and testis." Genomics Proteomics Bioinformatics **10**(2): 82-93.

D'Haene, B., C. Attanasio, et al. (2009). "Disease-causing 7.4 kb cis-regulatory deletion disrupting conserved non-coding sequences and their interaction with the FOXL2 promotor: implications for mutation screening." PLoS Genet **5**(6): e1000522.

Daggett, M. A., D. A. Rice, et al. (2000). "Expression of steroidogenic factor 1 in the testis requires an E box and CCAAT box in its promoter proximal region." Biol Reprod **62**(3): 670-679.

Dahl, J. A., A. H. Reiner, et al. (2010). "Histone H3 lysine 27 methylation asymmetry on developmentally-regulated promoters distinguish the first two lineages in mouse preimplantation embryos." PLoS One **5**(2): e9150.

Dankbar, B., M. H. Brinkworth, et al. (1995). "Ubiquitous expression of the androgen receptor and testis-specific expression of the FSH receptor in the cynomolgus monkey (Macaca fascicularis) revealed by a ribonuclease protection assay." J Steroid Biochem Mol Biol **55**(1): 35-41.

de Kok, Y. J., G. F. Merkx, et al. (1995). "A duplication/paracentric inversion associated with familial X-linked deafness (DFN3) suggests the presence of a regulatory element more than 400 kb upstream of the POU3F4 gene." Hum Mol Genet **4**(11): 2145-2150.

de la Calle-Mustienes, E., C. G. Feijoo, et al. (2005). "A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts." Genome Res **15**(8): 1061-1072.

Dermitzakis, E. T., A. Reymond, et al. (2005). "Conserved non-genic sequences - an unexpected feature of mammalian genomes." Nat Rev Genet **6**(2): 151-157.

Dermitzakis, E. T., A. Reymond, et al. (2002). "Numerous potentially functional but non-genic conserved sequences on human chromosome 21." Nature **420**(6915): 578-582.

Dias, J. A., B. D. Cohen, et al. (2002). "Molecular, structural, and cellular biology of follitropin and follitropin receptor." Vitam Horm **64**: 249-322.

Dickmeis, T. and F. Muller (2005). "The identification and functional characterisation of conserved regulatory elements in developmental genes." Brief Funct Genomic Proteomic **3**(4): 332-350.

Dorrington, J. H. and D. T. Armstrong (1979). "Effects of FSH on gonadal functions." Recent Prog Horm Res **35**: 301-342.

Dorrington, J. H., N. F. Roller, et al. (1974). "The effects of FSH on cell preparations from the rat testis." Curr Top Mol Endocrinol **1**: 237-241.

Drummond, A. E., M. Dyson, et al. (1996). "Differential responses of post-natal rat ovarian cells to FSH and activin." Mol Cell Endocrinol **122**(1): 21-32.

Dunkel, L., J. L. Tilly, et al. (1994). "Follicle-stimulating hormone receptor expression in the rat ovary: increases during prepubertal development and regulation by the opposing actions of transforming growth factors beta and alpha." Biol Reprod **50**(4): 940-948.

Edgar, R., M. Domrachev, et al. (2002). "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository." Nucleic Acids Res **30**(1): 207-210.

Engel, J. D. and K. Tanimoto (2000). "Looping, linking, and chromatin activity: new insights into beta-globin locus regulation." Cell **100**(5): 499-502.

Engstrom, P. G., S. J. Ho Sui, et al. (2007). "Genomic regulatory blocks underlie extensive microsynteny conservation in insects." Genome Res **17**(12): 1898-1908.

Epstein, D. J. (2009). "Cis-regulatory mutations in human disease." Brief Funct Genomic Proteomic **8**(4): 310-316.

Ewing, B. and P. Green (1998). "Base-calling of automated sequencer traces using phred. II. Error probabilities." Genome Res **8**(3): 186-194.

Ewing, B., L. Hillier, et al. (1998). "Base-calling of automated sequencer traces using phred. I. Accuracy assessment." Genome Res **8**(3): 175-185.

Feschotte, C. (2008). "Transposable elements and the evolution of regulatory networks." Nat Rev Genet **9**(5): 397-405.

Findlay, J. K. and A. E. Drummond (1999). "Regulation of the FSH Receptor in the Ovary." Trends Endocrinol Metab **10**(5): 183-188.

Fletcher, P. W. and L. E. Reichert, Jr. (1984). "Cellular processing of follicle-stimulating hormone by Sertoli cells in serum-free culture." Mol Cell Endocrinol **34**(1): 39-49.

Flores, J. A., J. D. Veldhuis, et al. (1990). "Follicle-stimulating hormone evokes an increase in intracellular free calcium ion concentrations in single ovarian (granulosa) cells." Endocrinology **127**(6): 3172-3179.

Fong, T. C. and B. M. Emerson (1992). "The erythroid-specific protein cGATA-1 mediates distal enhancer activity through a specialized beta-globin TATA box." Genes Dev **6**(4): 521-532.

Fromm, G. and M. Bulger (2009). "A spectrum of gene regulatory phenomena at mammalian beta-globin gene loci." Biochem Cell Biol **87**(5): 781-790.

Gibbs, R. A., G. M. Weinstock, et al. (2004). "Genome sequence of the Brown Norway rat yields insights into mammalian evolution." Nature **428**(6982): 493-521.

Giresi, P. G., J. Kim, et al. (2007). "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin." Genome Res **17**(6): 877-885.

Glazko, G. V., E. V. Koonin, et al. (2003). "A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions." Trends Genet **19**(3): 119-124.

Goetz, T. L., T. L. Lloyd, et al. (1996). "Role of E box and initiator region in the expression of the rat follicle-stimulating hormone receptor." J Biol Chem **271**(52): 33317-33324.

Gong, X. and E. A. McGee (2009). "Smad3 Is Required for Normal Follicular Follicle-Stimulating Hormone Responsiveness in the Mouse." Biol Reprod.

Gore-Langton, R. E. and J. H. Dorrington (1981). "FSH induction of aromatase in cultured rat granulosa cells measured by a radiometric assay." Mol Cell Endocrinol **22**(2): 135-151.

Grass, J. A., M. E. Boyer, et al. (2003). "GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling." Proc Natl Acad Sci U S A **100**(15): 8811-8816.

Grasso, P. and L. E. Reichert, Jr. (1990). "Follicle-stimulating hormone receptor-mediated uptake of 45Ca2+ by cultured rat Sertoli cells does not require activation of cholera toxin- or pertussis toxin-sensitive guanine nucleotide binding proteins or adenylate cyclase." Endocrinology **127**(2): 949-956.

Griswold, M. D. and J. S. Kim (2001). "Site-specific methylation of the promoter alters deoxyribonucleic acid-protein interactions and prevents follicle-stimulating hormone receptor gene transcription." Biol Reprod **64**(2): 602-610.

Gromoll, J., B. Dankbar, et al. (1994). "Characterization of the 5' flanking region of the human follicle-stimulating hormone receptor gene." Mol Cell Endocrinol **102**(1-2): 93-102.

Gromoll, J., T. Ried, et al. (1994). "Localization of the human FSH receptor to chromosome 2 p21 using a genomic probe comprising exon 10." J Mol Endocrinol **12**(3): 265-271.

Guo, Y., K. Monahan, et al. (2012). "CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice." Proc Natl Acad Sci U S A **109**(51): 21081-21086.

Hammond, S., J. C. Swanberg, et al. (2011). "Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology." BMC Genomics **12**: 67.

Handel, M. A. and J. J. Eppig (1979). "Sertoli cell differentiation in the testes of mice genetically deficient in germ cells." Biol Reprod **20**(5): 1031-1038.

Hardison, R. C. (2012). "Genome-wide epigenetic data facilitate understanding of disease susceptibility association studies." J Biol Chem **287**(37): 30932-30940.

Havlak, P., R. Chen, et al. (2004). "The Atlas genome assembly system." Genome Res **14**(4): 721-732.

Hawkins, R. D. and B. Ren (2006). "Genome-wide location analysis: insights on transcriptional regulation." Hum Mol Genet **15 Spec No 1**: R1-7.

Heckert, L. and M. D. Griswold (1993). "Expression of the FSH receptor in the testis." Recent Prog Horm Res **48**: 61-77.

Heckert, L. L. (2001). "Activation of the rat follicle-stimulating hormone receptor promoter by steroidogenic factor 1 is blocked by protein kinase a and requires upstream stimulatory factor binding to a proximal E box element." Mol Endocrinol **15**(5): 704-715.

Heckert, L. L. (2005). Structure and Regulation of the FSH Receptor Gene. Sertoli Cell Biology. M. K. a. G. Skinner, M.D.**:** 281-299.

Heckert, L. L., M. A. Daggett, et al. (1998). "Multiple promoter elements contribute to activity of the follicle-stimulating hormone receptor (FSHR) gene in testicular Sertoli cells." Mol Endocrinol **12**(10): 1499-1512.

Heckert, L. L., I. J. Daley, et al. (1992). "Structural organization of the follicle-stimulating hormone receptor gene." Mol Endocrinol **6**(1): 70-80.

Heckert, L. L. and M. D. Griswold (1991). "Expression of follicle-stimulating hormone receptor mRNA in rat testes and Sertoli cells." Mol Endocrinol **5**(5): 670-677.

Heckert, L. L. and M. D. Griswold (2002). "The expression of the follicle-stimulating hormone receptor in spermatogenesis." Recent Prog Horm Res **57**: 129-148.

Heckert, L. L., M. Sawadogo, et al. (2000). "The USF proteins regulate transcription of the follicle-stimulating hormone receptor but are insufficient for cell-specific expression." Mol Endocrinol **14**(11): 1836-1848.

Heindel, J. J., R. Rothenberg, et al. (1975). "LH and FSH stimulation of cyclic AMP in specific cell types isolated from the testes." J Cyclic Nucleotide Res **1**(2): 69-79.

Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.

Hermann, B. P. and L. L. Heckert (2005). "Silencing of Fshr occurs through a conserved, hypersensitive site in the first intron." Mol Endocrinol **19**(8): 2112-2131.

Hermann, B. P. and L. L. Heckert (2007). "Transcriptional regulation of the FSH receptor: new perspectives." Mol Cell Endocrinol **260-262**: 100-108.

Hermann, B. P., K. Hornbaker, et al. (2008). "In vivo regulation of follicle-stimulating hormone receptor by the transcription factors upstream stimulatory factor 1 and upstream stimulatory factor 2 is cell specific." Endocrinology **149**(10): 5297-5306.

Hermann, B. P., K. I. Hornbaker, et al. (2007). "Distal regulatory elements are required for Fshr expression, in vivo." Mol Cell Endocrinol **260-262**: 49-58.

Hiroi, H., L. K. Christenson, et al. (2004). "Regulation of transcription of the steroidogenic acute regulatory protein (StAR) gene: temporal and spatial changes in transcription factor binding and histone modification." Mol Cell Endocrinol **215**(1-2): 119-126.

Huhtaniemi, I. T., V. Eskola, et al. (1992). "The murine luteinizing hormone and follicle-stimulating hormone receptor genes: transcription initiation sites, putative promoter sequences and promoter activity." Mol Cell Endocrinol **88**(1-3): 55-66.

Huhtaniemi, I. T. and A. P. Themmen (2005). "Mutations in human gonadotropin and gonadotropin-receptor genes." Endocrine **26**(3): 207-217.

Jahnsen, T., J. O. Gordeladze, et al. (1980). "FSH-response adenylyl cyclase in rat testes: desensitization by homologous hormone." Arch Androl **5**(2): 169-177.

Jamieson, R. V., R. Perveen, et al. (2002). "Domain disruption and mutation of the bZIP transcription factor, MAF, associated with cataract, ocular anterior segment dysgenesis and coloboma." Hum Mol Genet **11**(1): 33-42.

Ji, Q., P. I. Liu, et al. (2004). "Follicle stimulating hormone-induced growth promotion and gene expression profiles on ovarian surface epithelial cells." Int J Cancer **112**(5): 803-814.

Jullien, N. and J. P. Herman (2011). "LUEGO: a cost and time saving gel shift procedure." Biotechniques **51**(4): 267-269.

Karl, A. F. and M. D. Griswold (1990). "Sertoli cells of the testis: preparation of cell cultures and effects of retinoids." Methods Enzymol **190**: 71-75.

Ketelslegers, J. M. and K. J. Catt (1974). "Receptor binding properties of 125I-hFSH prepared by enzymatic iodination." J Clin Endocrinol Metab **39**(6): 1159-1162.

Kidwell, M. G. and D. R. Lisch (2000). "Transposable elements and host genome evolution." Trends Ecol Evol **15**(3): 95-99.

Kikuta, H., M. Laplante, et al. (2007). "Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates." Genome Res **17**(5): 545-555.

Kiltie, A. E. (2010). "Common predisposition alleles for moderately common cancers: bladder cancer." Curr Opin Genet Dev.

Kim, J. S. and M. D. Griswold (2001). "E2F and GATA-1 are required for the Sertoli cell-specific promoter activity of the follicle-stimulating hormone receptor gene." J Androl **22**(4): 629-639.

Kim, T. H., L. O. Barrera, et al. (2005). "A high-resolution map of active promoters in the human genome." Nature **436**(7052): 876-880.

King, D. C., J. Taylor, et al. (2005). "Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences." Genome Res **15**(8): 1051-1060.

King, D. C., J. Taylor, et al. (2007). "Finding cis-regulatory elements using comparative genomics: some lessons from ENCODE data." Genome Res **17**(6): 775-786.

Kishi, H., T. Minegishi, et al. (1998). "The effect of activin and FSH on the differentiation of rat granulosa cells." FEBS Lett **422**(2): 274-278.

Kliesch, S., T. L. Penttila, et al. (1992). "Fsh Receptor Messenger-Rna Is Expressed Stage-Dependently during Rat Spermatogenesis." Molecular and Cellular Endocrinology **84**(3): R45-R49.

Knecht, M., T. Ranta, et al. (1983). "Granulosa cell differentiation in vitro: induction and maintenance of follicle-stimulating hormone receptors by adenosine 3',5'-monophosphate." Endocrinology **113**(3): 949-956.

Kolbe, D., J. Taylor, et al. (2004). "Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat." Genome Res **14**(4): 700-707.

Koo, Y. B., I. Ji, et al. (1991). "Structure of the luteinizing hormone receptor gene and multiple exons of the coding sequence." Endocrinology **128**(5): 2297-2308.

Koonin, E. V. (2003). "Comparative genomics, minimal gene-sets and the last universal common ancestor." Nat Rev Microbiol **1**(2): 127-136.

Koschmieder, S., F. Rosenbauer, et al. (2005). "Role of transcription factors C/EBPalpha and PU.1 in normal hematopoiesis and leukemia." Int J Hematol **81**(5): 368-377.

Kumar, T. R., Y. Wang, et al. (1997). "Follicle stimulating hormone is required for ovarian follicle maturation but not male fertility." Nat Genet **15**(2): 201-204.

Lagerstrom, M. C. and H. B. Schioth (2008). "Structural diversity of G protein-coupled receptors and significance for drug discovery." Nature Reviews Drug Discovery **7**(4): 339-357.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Leeson, C. R. and D. E. Forman (1981). "Postnatal development and differentiation of contractile cells within the rabbit testis." J Anat **132**(Pt 4): 491-511.

Lei, N. and L. L. Heckert (2002). "Sp1 and Egr1 regulate transcription of the Dmrt1 gene in Sertoli cells." Biol Reprod **66**(3): 675-684.

Levallet, J., P. Koskimies, et al. (2001). "The promoter of murine follicle-stimulating hormone receptor: functional characterization and regulation by transcription factor steroidogenic factor 1." Mol Endocrinol **15**(1): 80-92.

Li, J. W., R. Schmieder, et al. (2012). "SEQanswers: an open access community for collaboratively decoding genomes." Bioinformatics **28**(9): 1272-1273.

Li, X., X. Wang, et al. (2008). "High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression." Plant Cell **20**(2): 259-276.

Linder, C., L. Heckert, et al. (1994). "Follicle-stimulating hormone receptor gene promoter activity." Endocrine **2**: 957-966.

Linder, C., L. Heckert, et al. (1994). "Follicle-stimulating hormone receptor gene promoter activity." Endocrine **2**: 957-966.

Linder, C. C., L. L. Heckert, et al. (1994). "Follicle-Stimulating-Hormone Receptor Gene Promoter Activity." Endocrine **2**(10): 957-966.

Liska, F., P. Snajdr, et al. (2009). "Deletion of a conserved noncoding sequence in Plzf intron leads to Plzf down-regulation in limb bud and polydactyly in the rat." Dev Dyn **238**(3): 673-684.

Liu, Z. and W. T. Garrard (2005). "Long-range interactions between three transcriptional enhancers, active Vkappa gene promoters, and a 3' boundary sequence spanning 46 kilobases." Mol Cell Biol **25**(8): 3220-3231.

Lodder, E. M., B. H. Eussen, et al. (2009). "Implication of long-distance regulation of the HOXA cluster in a patient with postaxial polydactyly." Chromosome Res **17**(6): 737-744.

Lu, C. L., W. Yang, et al. (2009). "Inhibin A inhibits follicle-stimulating hormone (FSH) action by suppressing its receptor expression in cultured rat granulosa cells." Molecular and Cellular Endocrinology **298**(1-2): 48-56.

Maeda, K., S. Taniuchi, et al. (2012). "Pit-1w may regulate prolactin gene expression in mouse testis." Gen Comp Endocrinol **178**(2): 180-184.

Margulies, E. H., G. M. Cooper, et al. (2007). "Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome." Genome Res **17**(6): 760-774.

Mariani, S., L. Salvatori, et al. (2006). "Expression and cellular localization of follicle-stimulating hormone receptor in normal human prostate, benign prostatic hyperplasia and prostate cancer." J Urol **175**(6): 2072-2077; discussion 2077.

Martinez-Jimenez, C. P., M. J. Gomez-Lechon, et al. (2005). "Transcriptional regulation of the human hepatic CYP3A4: identification of a new distal enhancer region responsive to CCAAT/enhancer-binding protein beta isoforms (liver activating protein and liver inhibitory protein)." Mol Pharmacol **67**(6): 2088-2101.

McCann, J. A., E. M. Muro, et al. (2007). "ChIP on SNP-chip for genome-wide analysis of human histone H4 hyperacetylation." BMC Genomics **8**: 322.

McEwen, K. R. and A. C. Ferguson-Smith (2010). "Distinguishing epigenetic marks of developmental and imprinting regulation." Epigenetics Chromatin **3**(1): 2.

McGaughey, D. M., Z. E. Stine, et al. (2009). "Asymmetrical distribution of non-conserved regulatory sequences at PHOX2B is reflected at the ENCODE loci and illuminates a possible genome-wide trend." BMC Genomics **10**: 8.

McLean, D. J., P. J. Friel, et al. (2002). "Oligonucleotide microarray analysis of gene expression in follicle-stimulating hormone-treated rat Sertoli cells." Mol Endocrinol **16**(12): 2780-2792.

McNeilly, A. S., C. J. Souza, et al. (2002). "Production of inhibin A not B in rams: changes in plasma inhibin A during testis growth, and expression of inhibin/activin subunit mRNA and protein in adult testis." Reproduction **123**(6): 827-835.

Meachem, S. J., R. I. McLachlan, et al. (1996). "Neonatal exposure of rats to recombinant follicle stimulating hormone increases adult Sertoli and spermatogenic cell numbers." Biol Reprod **54**(1): 36-44.

Meachem, S. J., S. M. Ruwanpura, et al. (2005). "Developmentally distinct in vivo effects of FSH on proliferation and apoptosis during testis maturation." J Endocrinol **186**(3): 429-446.

Means, A. R. and C. Huckins (1974). "Coupled events in the early biochemical actions of FSH on the Sertoli cells of the testis." Curr Top Mol Endocrinol **1**: 145-165.

Minegishi, T., M. Tano, et al. (1995). "Regulation of follicle-stimulating hormone receptor messenger ribonucleic acid levels in cultured rat granulosa cells." Mol Cell Endocrinol **108**(1-2): 67-73.

Moguilevsky, J. A., C. Libertun, et al. (1970). "Metabolic sensitivity of different hypothalamic areas to luteinizing hormone (LH), follicle stimulating hormone (FSH), and testosterone." Neuroendocrinology **6**(3): 153-159.

Monaco, L., N. S. Foulkes, et al. (1995). "Pituitary follicle-stimulating hormone (FSH) induces CREM gene expression in Sertoli cells: involvement in long-term desensitization of the FSH receptor." Proc Natl Acad Sci U S A **92**(23): 10673-10677.

Monsees, G. M., P. Kraft, et al. (2011). "Comprehensive screen of genetic variation in DNA repair pathway genes and postmenopausal breast cancer risk." Breast Cancer Res Treat **125**(1): 207-214.

Montgomery, G. W., M. L. Tate, et al. (1995). "The follicle-stimulating hormone receptor and luteinizing hormone receptor genes are closely linked in sheep and deer." J Mol Endocrinol **15**(3): 259-265.

Mural, R. J., M. D. Adams, et al. (2002). "A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome." Science **296**(5573): 1661-1671.

Murphy, H. D. (1965). "Sertoli Cell Stimulation Following Intratesticular Injections of Fsh in the Hypophysectomized Rat." Proc Soc Exp Biol Med **118**: 1202-1205.

Nakamura, M., T. Minegishi, et al. (1993). "Effect of an Activin-a on Follicle-Stimulating-Hormone (Fsh) Receptor Messenger-Ribonucleic-Acid Levels and Fsh Receptor Expressions in Cultured Rat Granulosa-Cells." Endocrinology **133**(2): 538-544.

Nakatani, A., S. Shimasaki, et al. (1991). "Cyclic changes in follistatin messenger ribonucleic acid and its protein in the rat ovary during the estrous cycle." Endocrinology **129**(2): 603-611.

Nimrod, A., G. F. Erickson, et al. (1976). "A specific FSH receptor in rat granulosa cells: properties of binding in vitro." Endocrinology **98**(1): 56-64.

Nobrega, M. A., I. Ovcharenko, et al. (2003). "Scanning human gene deserts for long-range enhancers." Science **302**(5644): 413.

Noce, T., Y. Fujiwara, et al. (1993). "A novel murine zinc finger gene mapped within the tw18 deletion region expresses in germ cells and embryonic nervous system." Dev Biol **155**(2): 409-422.

Nordhoff, V., J. Gromoll, et al. (2003). "Targeted expression of human FSH receptor Asp567Gly mutant mRNA in testis of transgenic mice: role of human FSH receptor promoter." Asian J Androl **5**(4): 267-275.

O'Neill, L. P. and B. M. Turner (1995). "Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner." EMBO J **14**(16): 3946-3957.

O'Shaughnessy, P. J., P. Marsh, et al. (1994). "Follicle-stimulating hormone receptor mRNA in the mouse ovary during post-natal development in the normal mouse and in the adult hypogonadal (hpg) mouse: structure of alternate transcripts." Mol Cell Endocrinol **101**(1-2): 197-201.

Obara, N., N. Suzuki, et al. (2008). "Repression via the GATA box is essential for tissue-specific erythropoietin gene expression." Blood **111**(10): 5223-5232.

Orth, J. and A. K. Christensen (1977). "Localization of 125I-labeled FSH in the testes of hypophy-sectomized rats by autoradiography at the light and electron microscope levels." Endocrinology **101**(1): 262-278.

Orth, J. M. (1984). "The role of follicle-stimulating hormone in controlling Sertoli cell proliferation in testes of fetal rats." Endocrinology **115**(4): 1248-1255.

Orth, J. M., G. L. Gunsalus, et al. (1988). "Evidence from Sertoli cell-depleted rats indicates that spermatid number in adults depends on numbers of Sertoli cells produced during perinatal development." Endocrinology **122**(3): 787-794.

Orth, J. M., C. A. Higginbotham, et al. (1984). "Hemicastration causes and testosterone prevents enhanced uptake of [3H] thymidine by Sertoli cells in testes of immature rats." Biol Reprod **30**(1): 263-270.

Oshaughnessy, P. J., P. Marsh, et al. (1994). "Follicle-Stimulating-Hormone Receptor Messenger-Rna in the Mouse Ovary during Postnatal-Development in the Normal Mouse and in the Adult Hypogonadal (Hpg) Mouse - Structure of Alternate Transcripts." Molecular and Cellular Endocrinology **101**(1-2): 197-201.

Ovcharenko, I., G. G. Loots, et al. (2005). "Evolution and functional classification of vertebrate gene deserts." Genome Res **15**(1): 137-145.

Pauler, F. M., M. A. Sloane, et al. (2009). "H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome." Genome Res **19**(2): 221-233.

Pekowska, A., T. Benoukraf, et al. (2011). "H3K4 tri-methylation provides an epigenetic signature of active enhancers." EMBO J **30**(20): 4198-4210.

Peluso, J. J. and R. W. Steger (1978). "Role of FSH in regulating granulosa cell division and follicular atresia in rats." J Reprod Fertil **54**(2): 275-278.

Pennacchio, L. A. (2003). "Insights from human/mouse genome comparisons." Mamm Genome **14**(7): 429-436.

Pennacchio, L. A., N. Ahituv, et al. (2006). "In vivo enhancer analysis of human conserved non-coding sequences." Nature **444**(7118): 499-502.

Perez-Solis, M. A., H. Macias, et al. (2009). "Molecular cloning and functional analysis of the FSH receptor gene promoter from the volcano mouse (Neotomodon alstoni alstoni)." Endocrine.

Peterson, K. R., H. Fedosyuk, et al. (2012). "LCR 5' hypersensitive site specificity for globin gene activation within the active chromatin hub." Nucleic Acids Res **40**(22): 11256-11269.

Plant, T. M. (2008). "Hypothalamic control of the pituitary-gonadal axis in higher primates: key advances over the last two decades." J Neuroendocrinol **20**(6): 719-726.

Pollard, K. S., M. J. Hubisz, et al. (2010). "Detection of nonneutral substitution rates on mammalian phylogenies." Genome Res **20**(1): 110-121.

Prabhakar, S., F. Poulin, et al. (2006). "Close sequence comparisons are sufficient to identify human cis-regulatory elements." Genome Res **16**(7): 855-863.

Prior, J. C. (2007). "FSH and bone--important physiology or not?" Trends Mol Med **13**(1): 1-3.

Putowski, L. T., W. J. Schillings, et al. (2004). "Human follicle-stimulating hormone receptor (FSH-R) promoter/enhancer activity is inhibited by transcriptional factors, from the upstream stimulating factors family, via E-box and newly identified initiator element (Inr) in FSH-R non-expressing cells." Gynecol Endocrinol **19**(1): 9-17.

Radman-Livaja, M., C. L. Liu, et al. (2010). "Replication and active demethylation represent partially overlapping mechanisms for erasure of H3K4me3 in budding yeast." PLoS Genet **6**(2): e1000837.

Ramsey, S. A., T. A. Knijnenburg, et al. (2010). "Genome-wide histone acetylation data improve prediction of mammalian transcription factor binding sites." Bioinformatics **26**(17): 2071-2075.

Rando, O. J. (2012). "Combinatorial complexity in chromatin structure and function: revisiting the histone code." Curr Opin Genet Dev **22**(2): 148-155.

Rannikki, A. S., F. P. Zhang, et al. (1995). "Ontogeny of follicle-stimulating hormone receptor gene expression in the rat testis and ovary." Mol Cell Endocrinol **107**(2): 199-208.

Rannikko, A., T. L. Penttila, et al. (1996). "Stage-specific expression of the FSH receptor gene in the prepubertal and adult rat seminiferous epithelium." J Endocrinol **151**(1): 29-35.

Raychoudhury, S. S., E. W. Thompson, et al. (1993). "Sertoli cells as paracrine modulators of DNA synthesis in rat peritubular myoid cells in culture." J Reprod Fertil **99**(2): 513-518.

Rhead, B., D. Karolchik, et al. (2010). "The UCSC Genome Browser database: update 2010." Nucleic Acids Res **38**(Database issue): D613-619.

Richards, J. S., J. J. Ireland, et al. (1976). "Ovarian follicular development in the rat: hormone receptor regulation by estradiol, follicle stimulating hormone and luteinizing hormone." Endocrinology **99**(6): 1562-1570.

Rijnkels, M., E. Kabotyanski, et al. (2010). "The epigenetic landscape of mammary gland development and functional differentiation." J Mammary Gland Biol Neoplasia **15**(1): 85-100.

Ringrose, L., S. Chabanis, et al. (1999). "Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances." EMBO J **18**(23): 6630-6641.

Ritter, V., B. Thuering, et al. (2008). "Follicle-stimulating hormone does not impact male bone mass in vivo or human male osteoclasts in vitro." Calcif Tissue Int **82**(5): 383-391.

Rizzolio, F., S. Bione, et al. (2008). "Highly conserved non-coding sequences and the 18q critical region for short stature: a common mechanism of disease?" PLoS One **3**(1): e1460.

Robinson, L. J., I. Tourkova, et al. (2010). "FSH-Receptor Isoforms and FSH-dependent Gene Transcription in Human Monocytes and Osteoclasts." Biochem Biophys Res Commun.

Rodriguez, C. I., N. Girones, et al. (2003). "Cha, a basic helix-loop-helix transcription factor involved in the regulation of upstream stimulatory factor activity." J Biol Chem **278**(44): 43135-43145.

Roh, T. Y., W. C. Ngau, et al. (2004). "High-resolution genome-wide mapping of histone modifications." Nat Biotechnol **22**(8): 1013-1016.

Roh, T. Y., G. Wei, et al. (2007). "Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns." Genome Res **17**(1): 74-81.

Rousseau-Merck, M. F., M. Atger, et al. (1993). "The chromosomal localization of the human follicle-stimulating hormone receptor gene (FSHR) on 2p21-p16 is similar to that of the luteinizing hormone receptor gene." Genomics **15**(1): 222-224.

Ruh, M. F., R. Dunn, 2nd, et al. (1996). "Interrelationships between nuclear structure and ligand-activated intracellular receptors." Crit Rev Eukaryot Gene Expr **6**(2-3): 271-283.

Russell, L. D., L. R. de Franca, et al. (1995). "Characteristics of mitotic cells in developing and adult testes with observations on cell lineages." Tissue Cell **27**(1): 105-128.

Russell, L. D. and M. D. Griswold (1993). The Sertoli Cell. Clearwater, FL, Cache River Press.

Ruwanpura, S. M., R. I. McLachlan, et al. (2008). "Follicle-stimulating hormone affects spermatogonial survival by regulating the intrinsic apoptotic pathway in adult rats." Biol Reprod **78**(4): 705-713.

Rye, M., P. Saetrom, et al. (2011). "Clustered ChIP-Seq-defined transcription factor binding sites and histone modifications map distinct classes of regulatory elements." BMC Biol **9**: 80.

Ryser, S., D. Glauser, et al. (2011). "Gene expression profiling of rat spermatogonia and Sertoli cells reveals signaling pathways from stem cells to niche and testicular cancer cells to surrounding stroma." BMC Genomics **12**: 29.

Sairam, M. R., L. G. Jiang, et al. (1996). "Follitropin signal transduction: alternative splicing of the FSH receptor gene produces a dominant negative form of receptor which inhibits hormone action." Biochem Biophys Res Commun **226**(3): 717-722.

Sairam, M. R. and V. S. Subbarayan (1997). "Characterization of the 5' flanking region and potential control elements of the ovine follitropin receptor gene." Mol Reprod Dev **48**(4): 480-487.

Sanford, J. C. and B. E. Batten (1989). "Endocytosis of follicle-stimulating hormone by ovarian granulosa cells: analysis of hormone processing and receptor dynamics." J Cell Physiol **138**(1): 154-164.

Schmid, C. D. and P. Bucher (2007). "ChIP-Seq data reveal nucleosome architecture of human promoters." Cell **131**(5): 831-832; author reply 832-833.

Schneider, R. and R. Grosschedl (2007). "Dynamics and interplay of nuclear architecture, genome organization, and gene expression." Genes Dev **21**(23): 3027-3043.

Schoenborn, J. R., M. O. Dorschner, et al. (2007). "Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma." Nat Immunol **8**(7): 732-742.

Schwartz, N. B. (1982). "Role of ovarian inhibin (folliculostatin) in regulating FSH secretion in the female rat." Adv Exp Med Biol **147**: 15-36.

Seger, R., T. Hanoch, et al. (2001). "The ERK signaling cascade inhibits gonadotropin-stimulated steroidogenesis." J Biol Chem **276**(17): 13957-13964.

Shahmanesh, M., M. Sedigh, et al. (1980). "Feedback control of FSH secretion in the male rat." Horm Res **12**(5): 266-276.

Shetty, J., G. K. Marathe, et al. (1996). "Specific immunoneutralization of FSH leads to apoptotic cell death of the pachytene spermatocytes and spermatogonial cells in the rat." Endocrinology **137**(5): 2179-2182.

Shieh, B. H., R. S. Sparkes, et al. (1993). "Localization of the gene-encoding upstream stimulatory factor (USF) to human chromosome 1q22-q23." Genomics **16**(1): 266-268.

Shimizu, A., K. Tsutsui, et al. (1987). "Autoradiographic study of binding and internalization of follicle-stimulating hormone in the mouse testis minces in vitro." Endocrinol Jpn **34**(3): 431-442.

Shu, W., H. Chen, et al. (2011). "Genome-wide analysis of the relationships between DNaseI HS, histone modifications and gene expression reveals distinct modes of chromatin domains." Nucleic Acids Res **39**(17): 7428-7443.

Siepel, A., G. Bejerano, et al. (2005). "Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes." Genome Res **15**(8): 1034-1050.

Simoni, M., J. Gromoll, et al. (1997). "Molecular pathophysiology of the pituitary-gonadal axis." Adv Exp Med Biol **424**: 89-97.

Simoni, M., J. Gromoll, et al. (1997). "The follicle-stimulating hormone receptor: biochemistry, molecular biology, physiology, and pathophysiology." Endocr Rev **18**(6): 739-773.

Singh, J. and D. J. Handelsman (1996). "Neonatal administration of FSH increases Sertoli cell numbers and spermatogenesis in gonadotropin-deficient (hpg) mice." J Endocrinol **151**(1): 37-48.

Sirito, M., Q. Lin, et al. (1994). "Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells." Nucleic Acids Res **22**(3): 427-433.

Sites, C. K., K. Patterson, et al. (1994). "Follicle-Stimulating-Hormone (Fsh) Increases Fsh Receptor Messenger-Ribonucleic-Acid While Decreasing Fsh Binding in Cultured Porcine Granulosa-Cells." Endocrinology **134**(1): 411-417.

Smeenk, G., A. J. de Groot, et al. (2010). "Rad51C is essential for embryonic development and haploinsufficiency causes increased DNA damage sensitivity and genomic instability." Mutat Res **689**(1-2): 50-58.

Sokka, T. and I. Huhtaniemi (1990). "Ontogeny of gonadotrophin receptors and gonadotrophin-stimulated cyclic AMP production in the neonatal rat ovary." J Endocrinol **127**(2): 297-303.

Spiteri-Grech, J., G. F. Weinbauer, et al. (1993). "Effects of FSH and testosterone on intratesticular insulin-like growth factor-I and specific germ cell populations in rats treated with gonadotrophin-releasing hormone antagonist." J Endocrinol **137**(1): 81-89.

Sprengel, R., T. Braun, et al. (1990). "The testicular receptor for follicle stimulating hormone: structure and functional expression of cloned cDNA." Mol Endocrinol **4**(4): 525-530.

Steinberger, A., J. J. Heindel, et al. (1975). "Isolation and culture of FSH responsive Sertoli cells." Endocr Res Commun **2**(3): 261-272.

Steinberger, A. and E. Steinberger (1971). "Replication pattern of Sertoli cells in maturing rat testis in vivo and in organ culture." Biol Reprod **4**(1): 84-87.

Steinberger, A. and E. Steinberger (1971). "Replication Pattern of Sertoli Cells in Maturing Rat Testis in Vivo and in Organ Culture." Biology of Reproduction **4**: 84-87.

Steinberger, A., K. H. Thanki, et al. (1974). "FSH binding in rat testes during maturation and following hypophysectomy. Cellular localization of FSH receptors." Curr Top Mol Endocrinol **1**: 177-191.

Strahle, U. and S. Rastegar (2008). "Conserved non-coding sequences and transcriptional regulation." Brain Res Bull **75**(2-4): 225-230.

Sun, L., Y. Peng, et al. (2006). "FSH directly regulates bone mass." Cell **125**(2): 247-260.

Sun, L., Z. Zhang, et al. (2010). "Further Evidence for Direct Pro-Resorptive Actions of FSH." Biochem Biophys Res Commun.

Taniuchi, S., K. Maeda, et al. (2011). "Identification of mammalian Pit-1w, possibly involved in spermatogenesis in mice." Gen Comp Endocrinol **173**(2): 289-294.

Tano, M., T. Minegishi, et al. (1997). "Transcriptional and post-transcriptional regulation of FSH receptor in rat granulosa cells by cyclic AMP and activin." J Endocrinol **153**(3): 465-473.

Tano, M., T. Minegishi, et al. (1995). "Regulation of follistatin messenger ribonucleic acid in cultured rat granulosa cells." Mol Cell Endocrinol **109**(2): 167-174.

Taylor, J., S. Tyekucheva, et al. (2006). "ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements." Genome Res **16**(12): 1596-1604.

Tena-Sempere, M., P. R. Manna, et al. (1999). "Molecular cloning of the mouse follicle-stimulating hormone receptor complementary deoxyribonucleic acid: functional expression of alternatively spliced variants and receptor inactivation by a C566T transition in exon 7 of the coding sequence." Biol Reprod **60**(6): 1515-1527.

Themmen, A. P., L. J. Blok, et al. (1991). "Follitropin receptor down-regulation involves a cAMP-dependent post-transcriptional decrease of receptor mRNA expression." Mol Cell Endocrinol **78**(3): R7-13.

Tian, Y., Z. Jia, et al. (2011). "Global mapping of H3K4me1 and H3K4me3 reveals the chromatin state-based cell type-specific gene regulation in human Treg cells." PLoS One **6**(11): e27770.

Tindall, D. J., W. T. Schrader, et al. (1974). "The production of androgen binding protein by Sertoli Cells." Curr Top Mol Endocrinol **1**: 167-175.

Tsai-Morris, C. H., E. Buczko, et al. (1991). "Structural organization of the rat luteinizing hormone (LH) receptor gene." J Biol Chem **266**(17): 11355-11359.

Tung, P. S., M. K. Skinner, et al. (1984). "Cooperativity between Sertoli cells and peritubular myoid cells in the formation of the basal lamina in the seminiferous tubule." Ann N Y Acad Sci **438**: 435-446.

Vakoc, C. R., D. L. Letting, et al. (2005). "Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1." Mol Cell **17**(3): 453-462.

Vassart, G., L. Pardo, et al. (2004). "A molecular dissection of the glycoprotein hormone receptors." Trends Biochem Sci **29**(3): 119-126.

Vazquez, B. N., T. Laguna, et al. (2009). "CD69 gene is differentially regulated in T and B cells by evolutionarily conserved promoter-distal elements." J Immunol **183**(10): 6513-6521.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.

Verhoeven, G., E. Hoeben, et al. (2000). "Peritubular cell-Sertoli cell interactions: factors involved in PmodS activity." Andrologia **32**(1): 42-45.

Viollet, B., A. M. Lefrancois-Martinez, et al. (1996). "Immunochemical characterization and transacting properties of upstream stimulatory factor isoforms." J Biol Chem **271**(3): 1405-1415.

Visel, A., M. J. Blow, et al. (2009). "ChIP-seq accurately predicts tissue-specific activity of enhancers." Nature **457**(7231): 854-858.

Visel, A., J. Bristow, et al. (2007). "Enhancer identification through comparative genomics." Semin Cell Dev Biol **18**(1): 140-152.

Visel, A., S. Prabhakar, et al. (2008). "Ultraconservation identifies a small subset of extremely constrained developmental enhancers." Nat Genet **40**(2): 158-160.

Visel, A., Y. Zhu, et al. (2010). "Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice." Nature **464**(7287): 409-412.

Viswanathan, P., M. A. Wood, et al. (2009). "Follicle-stimulating hormone (FSH) transiently blocks FSH receptor transcription by increasing inhibitor of deoxyribonucleic acid binding/differentiation-2 and decreasing upstream stimulatory factor expression in rat Sertoli cells." Endocrinology **150**(8): 3783-3791.

Wang, Q., J. S. Carroll, et al. (2005). "Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking." Mol Cell **19**(5): 631-642.

Ward, M. C., M. D. Wilson, et al. (2013). "Latent regulatory potential of human-specific repetitive elements." Mol Cell **49**(2): 262-272.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-562.

Wicks, K. and J. C. Knight (2011). "Transcriptional repression and DNA looping associated with a novel regulatory element in the final exon of the lymphotoxin-beta gene." Genes Immun **12**(2): 126-135.

Woodruff, T. K., J. D'Agostino, et al. (1988). "Dynamic changes in inhibin messenger RNAs in rat ovarian follicles during the reproductive cycle." Science **239**(4845): 1296-1299.

Woolfe, A., M. Goodson, et al. (2005). "Highly conserved non-coding sequences are associated with vertebrate development." PLoS Biol **3**(1): e7.

Wu, J., L. T. Smith, et al. (2006). "ChIP-chip comes of age for genome-wide functional analysis." Cancer Res **66**(14): 6899-6902.

Wunsch, A., Y. Ahda, et al. (2005). "Single-nucleotide polymorphisms in the promoter region influence the expression of the human follicle-stimulating hormone receptor." Fertil Steril **84**(2): 446-453.

Xing, W., N. Danilovich, et al. (2002). "Orphan receptor chicken ovalbumin upstream promoter transcription factors inhibit steroid factor-1, upstream stimulatory factor, and activator protein-1 activation of ovine follicle-stimulating hormone receptor expression via composite cis-elements." Biol Reprod **66**(6): 1656-1666.

Xing, W. and M. R. Sairam (2001). "Characterization of regulatory elements of ovine follicle-stimulating hormone (FSH) receptor gene: the role of E-box in the regulation of ovine FSHreceptor expression." Biol Reprod **64**(2): 579-589.

Xing, W. and M. R. Sairam (2002). "Cross talk of two Krupple transcription factors regulates expression of the ovine FSH receptor gene." Biochem Biophys Res Commun **295**(5): 1096-1101.

Xing, W. and M. R. Sairam (2002). "Retinoic acid mediates transcriptional repression of ovine follicle-stimulating hormone receptor gene via a pleiotropic nuclear receptor response element." Biol Reprod **67**(1): 204-211.

Xu, J., Z. Shao, et al. (2012). "Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis." Dev Cell **23**(4): 796-811.

Xu, X., K. Tsumagari, et al. (2009). "DNaseI hypersensitivity at gene-poor, FSH dystrophy-linked 4q35.2." Nucleic Acids Res **37**(22): 7381-7393.

Yen, S. S. and C. C. Tsai (1971). "The biphasic pattern in the feedback action of ethinyl estradiol on the release of pituitary FSH and LH." J Clin Endocrinol Metab **33**(6): 882-887.

Yin, H., S. Sweeney, et al. (2011). "A high-resolution whole-genome map of key chromatin modifications in the adult Drosophila melanogaster." PLoS Genet **7**(12): e1002380.

Young, E. A. (1995). "The role of gonadal steroids in hypothalamic-pituitary-adrenal axis regulation." Crit Rev Neurobiol **9**(4): 371-381.

Young, M. D., T. A. Willson, et al. (2011). "ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity." Nucleic Acids Res **39**(17): 7415-7427.

Yu, M., L. Riva, et al. (2009). "Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis." Mol Cell **36**(4): 682-695.

Zaidi, M., H. C. Blair, et al. (2009). "New insights: elevated follicle-stimulating hormone and bone loss during the menopausal transition." Curr Rheumatol Rep **11**(3): 191-195.

Zaidi, M., H. C. Blair, et al. (2007). "Proresorptive actions of FSH and bone loss." Ann N Y Acad Sci **1116**: 376-382.

Zaidi, S., L. L. Zhu, et al. (2007). "Regulation of FSH receptor promoter activation in the osteoclast." Biochem Biophys Res Commun **361**(4): 910-915.

Zang, C., D. E. Schones, et al. (2009). "A clustering approach for identification of enriched domains from histone modification ChIP-Seq data." Bioinformatics **25**(15): 1952-1958.

Zentner, G. E. and P. C. Scacheri (2012). "The chromatin fingerprint of gene enhancer elements." J Biol Chem **287**(37): 30888-30896.

Zhang, X. Y., J. Chen, et al. (2009). "Follicle-stimulating hormone peptide can facilitate paclitaxel nanoparticles to target ovarian carcinoma in vivo." Cancer Res **69**(16): 6506-6514.

Zheng, W., M. S. Magid, et al. (1996). "Follicle-stimulating hormone receptor is expressed in human ovarian surface epithelium and fallopian tube." Am J Pathol **148**(1): 47-53.

Zheng, Y., J. Zhang, et al. (2010). "Screening RAD51C nucleotide alterations in patients with a family history of breast and ovarian cancer." Breast Cancer Res Treat **124**(3): 857-861.

Zhou, G., J. Wang, et al. (2004). "Cloning, expression and subcellular localization of HN1 and HN1L genes, as well as characterization of their orthologs, defining an evolutionarily conserved gene family." Gene **331**: 115-123.